Chunking the input: on the role of frequency and prosody in the segmentation strategies of adult bilinguals

Irene de la Cruz Pavía

2012

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Dirigida por: Doctora Itziar Laka Mugarza y Doctor Gorka Elordieta Alcibar

Universidad del País Vasco (UPV/EHU)

Departamento de Lingüística y Estudios Vascos

Programa de Doctorado *Lingüística/Hizkuntzalaritza*

La investigación presentada en esta tesis doctoral ha sido financiada por la beca del Programa de Formación y Perfeccionamiento del Personal Investigador del Gobierno Vasco BFI07.253.

Contents

List of Tables

List of Graphs

List of Figures

Acknowledgements

It is a truth universally acknowledged, that a young researcher who has at last finished her dissertation, must be in great debt to a large number of people...

I would first like to thank my supervisors, Itziar Laka and Gorka Elordieta. I have learnt countless things from them. Not only have I acquired many skills and knowledge essential for a researcher and for the development of this dissertation—I now know that I should avoid starting every single paragraph with “thus”—, but I have also learnt some priceless life lessons from them.

During these years, I have had the help and input of many people, for which I’m very grateful: Manuel Carreiras and Jon Andoni Duñabeitia both during my short visit at La Laguna and subsequently; Doug Saddy and Kelly Sloan for the three great months I spent at the CINN in Reading and everybody I met there, specially Carla Contemori and Magda Capó; Inma Hernaez and the members of AHOLAB; Urtzi Etxeberria and Jasone Salaberria from IKER-CNRS and María Arana and Murielle Fraty from UIT Bayonne; Nuria Sebastián-Gallés, Judit Gervain and Juan Manuel Toro for their invaluable input; as well as all the participants who took part in the experiments.

I’m very thankful to my colleagues and friends at the Laboratory of Psycholinguistics and the Faculty of Arts of the University of the Basque Country. I’ve shared countless laughs with them and discussed and learnt plenty from all of them. Thank you (alphabetically): Yolanda Acedo, Joana Acha, Aitziber Agirre, Agurtzane Azkarai, Kepa Erdozia, Marijo Ezeizabarrena, Jorge González, Junkal Gutierrez, Anna Hatzidaki, Aritz Irurtzun, Ibon Manterola, Amaia Munarriz, Leticia Pablos, Luis Pastor, Edurne Petrirena, Idoia Ros, Mikel Santesteban, Ander Tarlowski, Oxel Uribe-Etxebarria, Izaskun Villareal, Julian Villegas, Ruth Yates, Iraia Yetano and Adam

Zawiszewski. I hope I'm not forgetting anyone…if I am, please excuse me, it's the pressure!

I'm specially grateful to Agurtzane Azkarai for her friendship during these years, Julian Villegas for reading this dissertation (he offered!) and paying attention to every detail, Mikel Santesteban also for his comments on this dissertation, Idoia Ros and Adam Zawiszewski for their priceless help running the experiments and Yolanda Acedo and Edurne Petrirena for their impeccable technical and administrative support. Yolanda, what would I have done without you? Last, I would like to thank Adam Zawiszewski and specially Kepa Erdozia (who had the misfortune of being at the lab from the beginning) for their great help during these years with my maaaany questions and doubts, some of them interesting, some of them not so much... They have been amazingly patient and always kind. Karma owes you big time!

I also wish to thank my sister Elena for agreeing to support me for life as soon as I declared my intentions to embark in a dissertation (I might take you up on that), Tultu for keeping me sane during the long and frustating hours while trying to write at home and Olatz Bilbao for accompanying me since the dawn of time...

Finally, I wish to dedicate this dissertation to my parents, Pedro and María Luisa. Not only have they always supported me but I have learnt and keep learning a great deal from them every day. I could not have done this dissertation without you. Thanks.

Abstract

The present dissertation investigates the abilities of adult monolingual and bilingual speakers to implement statistical and prosodic cues in speech segmentation. Three are the aims of the present dissertation: (1) to examine whether bilingual speakers deploy the prosodic and statistical segmentation strategies that characterize their two languages, (2) to investigate the role that statistical and prosodic cues play in adult speech segmentation, and (3) to explore whether adult speakers make use of two types of cues that have been proposed as potentially allowing infants to determine the basic word order patter (OV/VO, head-initial or head-final) of the language under acquisition: the frequency distribution of functors and content words in natural languages (frequency-based cue) and the relative prominence within phonological phrases (prosodic cue).

Three artificial language learning experiments were conducted, in which the segmentation preferences of ambiguous artificial languages that contain these frequency-based and prosodic cues by adult monolingual and bilingual speakers were examined.

The results of the experiments showed that (1) bilingual speakers are able to implement the frequency-based segmentation strategies that characterize their two languages, though acquisition of the L2's segmentation strategy appears to be constrained, (2) statistical and prosodic cues seem to be outranked by acoustic-phonetic cues, supporting thus a hierarchical account of segmentation cues in which statistical and prosodic cues are the least weighed by adult speakers, (3) frequent-initial segmentation might be the universally preferred segmentation strategy, (4) frequency-based segmentation strategies are available segmentation cues to adult speakers.

Abstract

La presente tesis doctoral investiga las habilidades de los hablantes adultos monolingües y bilingües para implementar pistas estadísticas y prosódicas en la segmentación del habla. Tres son los objetivos de la presente tesis doctoral: (1) examinar si los hablantes bilingües hacen uso de las estrategias de segmentación estadísticas y prosódicas que caracterizan sus dos lenguas, (2) investigar el papel que juegan las pistas prosódicas y estadísticas en la segmentación del habla en hablantes adultos, y (3) explorar si los hablantes adultos emplean dos tipos de pistas que, tal y como se ha propuesto, permitirían potencialmente a los niños determinar el patrón básico de orden de palabras (OV/VO, núcleo inicial o núcleo final) de la lengua que están adquiriendo: la distribución de la frecuencia de los elementos funcionales y de contenido en las lenguas naturales (pista basada en la frecuencia) y la prominencia relativa dentro de los sintagmas fonológicos (pista prosódica).

Se llevaron a cabo tres experimentos de aprendizaje de lenguas artificiales, en los cuales se examinaron las preferencias de segmentación de hablantes adultos monolingües y bilingües de lenguas artificiales ambiguas que contenían pistas basadas en la frecuencia y pistas prosódicas.

Los resultados de estos experimentos mostraron que: (1) los hablantes bilingües son capaces de implementar las estrategias de segmentación basadas en la frecuencia que caracterizan sus dos lenguas, aunque la adquisición de la estrategia de segmentación de la L2 parece estar limitada; (2) las pistas estadísticas y prosódicas parecen verse sobrepasadas por las pistas acústico-fonéticas, apoyando por tanto una concepción jerárquica de las pistas de segmentación en la cual las pistas estadísticas y prosódicas son aquéllas a las que los hablantes adultos otorgan menor peso; (3) la segmentación en la cual los elementos frecuentes ocupan una posición final podría ser la estrategia de segmentación preferida universalmente; (4) las estrategias de segmentación basadas en la frecuencia son pistas de segmentación disponibles para los hablantes adultos.

Resumen

Las palabras y los sintagmas no son elementos diferenciados en el habla. Las fronteras de palabras y sintagmas no están marcadas de forma sistemática por pistas tales como pausas. A pesar del carácter continuo del habla, la segmentación del habla es un proceso automático e inconsciente para los hablantes adultos y una tarea que los niños llevan a cabo aparentemente sin esfuerzo. La segmentación del habla resulta de la adquisición e integración de diferentes estrategias perceptuales a lo largo del desarrollo. En las últimas décadas, numerosos estudios han investigado las pistas presentes en el input que permiten a los humanos (tanto niños como adultos) segmentar el habla en palabras, cruciales para la formación del léxico, así como en constituyentes sintácticos mayores tales como los sintagmas, las cuales permiten a los niños dividir el input en unidades menores y detectar las regularidades presentes en el mismo, emprendiendo potencialmente la adquisición de reglas sintácticas.

Tanto adultos como niños emplean pistas prosódicas, estadísticas, fonotácticas y acústico-fonéticas tales como la variación alofónica y la coarticulación, para segmentar el habla en palabras y sintagmas. Emplear estas pistas con éxito requiere con frecuencia el conocimiento de propiedades específicas de la lengua nativa. Así, aunque la sensibilidad hacia las propiedades rítmicas de las lenguas parece emerger incluso antes del nacimiento, los niños comienzan a segmentar palabras que conforman el ritmo característico de la lengua nativa hacia los siete meses y medio de edad. Este conocimiento de propiedades específicas de las lenguas, esencial para poder emplear las pistas de segmentación del habla, revela el interés de investigar la habilidad de los hablantes bilingües para implementar las estrategias de segmentación que caracterizan sus dos lenguas.

Entre estas pistas de segmentación disponibles, dos parecen jugar un papel crucial en los estadios iniciales de la adquisición del lenguaje: las pistas prosódicas y

estadísticas. Estas dos pistas son las primeras empleadas por los niños para localizar las fronteras de las palabras y son el objeto de investigación de la presente tesis doctoral.

Como se presenta en el capítulo 2, sección 2.1 de la presente tesis doctoral, tanto los adultos como los niños son sensibles a diversos tipos de información estadística presentes en el input, información de la que se valen para segmentar el habla, construir una representación de la estructura del input y descubrir regularidades. Se ha propuesto que la información estadística permitiría a los niños comenzar a construir el léxico y adquirir las restricciones fonotácticas y las regularidades sintácticas que caracterizan la lengua que están adquiriendo. Valiéndose de esta información, los niños establecerían el valor de propiedades gramaticales básicas de la lengua, emprendiendo así la adquisición de la sintaxis.

La investigación realizada en las últimas décadas sobre las propiedades estadísticas y distribucionales de los elementos funcionales en las lenguas naturales ha revelado que estos elementos podrían ser especialmente importantes para emprender la adquisición de la sintaxis, permitiendo a los niños establecer los valores de parámetros sintácticos principales antes incluso de poseer conocimiento léxico. Las lenguas naturales cuentan con elementos funcionales y de contenido. Estas dos categorías se distinguen por una serie de propiedades fonológicas, distribucionales y estadísticas. Los elementos funcionales se caracterizan por ser breves fonológicamente, por su alta frecuencia de aparición y por su tendencia a ocupar los extremos de los sintagmas. Los elementos funcionales actúan como puntos de anclaje que señalan las construcciones sintácticas y facilitan la extracción de regularidades, como ha sido demostrado en estudios de aprendizaje de lenguas artificiales. Los niños son muy sensibles a la frecuencia de aparición de los elementos funcionales. Su alta frecuencia de aparición les permite segmentar no sólo los elementos funcionales del habla, sino también los elementos de contenido adyacentes, así como categorizar nuevas palabras en las categorías funcionales y léxicas y adquirir determinados aspectos de la estructura sintáctica.

Además de las mencionadas pistas estadísticas, tanto los niños como los adultos emplean las pistas prosódicas presentes en el input para segmentar el habla en palabras y localizar los límites de los sintagmas, como se presenta en el capítulo 2, sección 2.2 de la presente tesis doctoral. Así, los hablantes explotan las propiedades rítmicas de la lengua nativa para segmentar las palabras del habla continua. A pesar de que no existe una correlación exacta entre los constituyentes sintácticos y los prosódicos, los constituyentes prosódicos tienden a alinearse con las fronteras de constituyentes sintácticos principales. Los sintagmas entonativos y fonológicos (dos miembros de la jerarquía fonológica) tienden a alinearse con las fronteras de las cláusulas y los sintagmas. La investigación sobre las propiedades acústicas de las fronteras de los sintagmas y cláusulas ha revelado que estos límites están frecuentemente marcados por pausas, alargamiento final y cambios en la frecuencia fundamental y que estas pistas parecen ser más frecuentes y estar hiperarticuladas en el habla dirigida a los niños.

La sensibilidad hacia las pistas prosódicas que marcan los constituyentes sintácticos aparece en un estadio muy temprano del desarrollo. Entre los seis y nueve meses de edad los niños emplean pistas prosódicas que señalan las fronteras de las cláusulas y los sintagmas para segmentar el habla. Los adultos, al igual que los niños, son sensibles a los marcadores prosódicos de los constituyentes sintácticos. Así, las fronteras de los sintagmas fonológicos restringen el acceso léxico y facilitan la segmentación y el aprendizaje de la estructura de las lenguas artificiales.

Se ha propuesto que las pistas estadísticas y prosódicas podrían permitir a los niños establecer el valor del parámetro de direccionalidad de la cabeza sintáctica, construyendo así una representación rudimentaria del orden de palabras de la lengua que están adquiriendo y emprendiendo por tanto la adquisición de la sintaxis.

Como se presenta en el capítulo 3, sección 3.1 de la presente tesis doctoral, Gervain (2007) propuso la *frequency-based bootstrapping hypothesis*, según la cual una pista estadística, la distribución de la frecuencia y el orden relativo de los elementos funcionales y las palabras de contenido en las lenguas naturales, permitiría a los niños establecer el valor del parámetro de direccionalidad de la cabeza sintáctica antes de poseer conocimiento léxico. El orden de los elementos funcionales y los elementos de contenido se correlaciona con el orden de las cabezas y los complementos en las lenguas naturales: en las lenguas de cabeza inicial (VO) los elementos funcionales aparecen típicamente en posición inicial en los sintagmas, mientras que en las lenguas de cabeza final (OV) los elementos funcionales aparecen en posición final en los sintagmas. Esta hipótesis propone que computar el orden relativo de los elementos frecuentes e infrecuentes en la lengua que se está adquiriendo permitiría a los niños descubrir el valor del parámetro sintáctico.

Resumen

Gervain (2007) obtiene evidencia que apoya esta hipótesis en una serie de experimentos de aprendizaje de lenguas artificiales con niños de ocho meses de edad y hablantes adultos de lenguas de cabeza inicial y lenguas de cabeza final. Gervain (2007) encontró que tanto los hablantes adultos como los niños computaban la distribución de la frecuencia de los elementos de una lengua artificial ambigua caracterizada por la alternancia de elementos frecuentes e infrecuentes, y que los hablantes de lenguas de cabeza final (euskera, japonés) preferían una segmentación de la lengua artificial ambigua en la que los elementos frecuentes ocupaban una posición final mientras que los hablantes de lenguas de cabeza inicial (francés, italiano) preferían una segmentación en la cual los elementos frecuentes ocupaban una posición inicial. El orden relativo de los elementos frecuentes (elementos funcionales) e infrecuentes (elementos de contenido) en la lengua nativa parecía por tanto influir en las preferencias de segmentación de la lengua artificial de los hablantes adultos y niños.

Paralelamente, la *phonological bootstrapping hypothesis* propone que tanto los niños como los adultos emplean pistas fonológicas para segmentar el habla, localizando las fronteras de los constituyentes sintácticos y dividiendo así el habla en pedazos de menor tamaño, lo cual permitiría a los niños detectar regularidades sintácticas. Nespor, Guasti y Christophe (1996) proponen que una pista prosódica – la localización y realización de la prominencia dentro de los sintagmas fonológicos – permitiría a los niños establecer el valor del parámetro de direccionalidad de la cabeza sintáctica, dada la correlación existente entre la localización de la prominencia en los sintagmas fonológicos y el orden relativo de cabezas y complementos en las lenguas naturales. En las lenguas de cabeza inicial la prominencia recae en la palabra situada en el extremo derecho del sintagma fonológico (el complemento), mientras que en las lenguas de cabeza final la prominencia recae en la palabra situada en el extremo izquierdo del sintagma fonológico (por tanto también el complemento). Nespor, Shukla, van de Vijver, Avesani, Schraudolf y Donati (2008) proponen que la prominencia se realiza de forma diferente en las lenguas de cabeza final (OV) y cabeza inicial (VO) como consecuencia de la ley Yámbico/Trocaico: en las lenguas de cabeza inicial la prominencia se realiza mediante a un alargamiento final, mientras que en las lenguas de cabeza final la prominencia se realiza mediante a una subida tonal. Así, la realización acústica de la prominencia permitiría a los niños establecer el orden de cabezas y complementos en la lengua que están adquiriendo.

Además de las pistas estadísticas y prosódicas descritas antes, niños y adultos emplean pistas segmentales en la segmentación del habla, tales como la coarticulación, la variación alofónica y las restricciones fonotácticas, como se presenta en el capítulo 2 sección 2.3 de la presente tesis doctoral. Los hablantes adultos, además, se apoyan también en la información léxico-semántica para segmentar el habla. Numerosos estudios han investigado el peso que los hablantes otorgan a las distintas pistas de segmentación de las que disponen. Estos estudios han revelado que las pistas de segmentación parecen estar organizadas de forma jerárquica, dependiendo del peso otorgado a cada pista, tal y como proponen Mattys, White y Melhorn (2005), y que, a lo largo del desarrollo, el peso relativo que reciben estas pistas cambia.

Los adultos se apoyan principalmente en las pistas léxicas para la segmentación, aunque la segmentación está modulada por las otras pistas presentes en el input. Cuando el habla no contiene información léxica o ésta es ambigua, los hablantes se apoyan en pistas segmentales. Únicamente cuando estas pistas segmentales son también ininteligibles se apoyan los hablantes adultos en las pistas prosódicas y estadísticas, siendo las pistas prosódicas preferidas a las estadísticas. Por tanto, las pistas estadísticas parecen ser empleadas como último recurso. Por el contrario, durante los estadios iniciales del desarrollo los niños otorgan mayor peso a aquellas pistas que no requieren un conocimiento detallado de las propiedades de la lengua que están adquiriendo (pistas prosódicas y estadísticas) y, a lo largo del segundo semestre del primer año de vida, adquirieren paulatinamente las pistas segmentales.

Tres son los objetivos principales de la presente tesis doctoral:

(a) En primer lugar, esta tesis investiga las habilidades de los hablantes bilingües adultos para emplear las estrategias de segmentación prosódicas y estadísticas que caracterizan sus dos lenguas. Por tanto, esta tesis doctoral pretende contribuir al conocimiento existente sobre el procesamiento del lenguaje en bilingües.

(b) Esta tesis doctoral explora además el papel que las pistas estadísticas y prosódicas juegan en la segmentación del habla en adultos.

(c) Por último, esta tesis pretende examinar si los hablantes adultos pueden emplear dos tipos de información que, como se ha propuesto, podrían permitir a los niños descubrir el orden de palabras básico de la lengua que están adquiriendo y establecer el valor del parámetro de direccionalidad de la cabeza sintáctica: la distribución de la frecuencia de los elementos funcionales y elementos de contenido en las lenguas naturales (pista basada en la frecuencia) y la prominencia relativa dentro de los sintagmas fonológicos (pista prosódica).

Para ello se presentan los resultados de tres experimentos de aprendizaje de lenguas artificiales, en los cuales se examinan las preferencias de segmentación de lenguas artificiales ambiguas que contienen la mencionada pista de frecuencia y la mencionada pista prosódica por parte de hablantes adultos monolingües y bilingües.

Experimentos A, B y C

En el experimento A, presentado en el capítulo 3, se empleó la lengua artificial diseñada y empleada por Gervain (2007), lengua caracterizada por la alternancia estricta de tres categorías de elementos frecuentes (*a*, *b*, *c*) con una sílaba compuesta por una consonante y una vocal (CV) en cada categoría y tres categorías de elementos infrecuentes (*X*, *Y*, *Z*), con nueve sílabas CV en cada categoría. Las sílabas frecuentes aparecían por tanto con una frecuencia nueve veces mayor que las sílabas infrecuentes. La familiarización se construyó concatenando una unidad hexasilábica prosódicamente plana con la estructura *aXbYcZ*. La intensidad de los primeros 15 segundos de la cadena se aumentó gradualmente y la intensidad de los 15 segundos finales se disminuyó gradualmente, creando una cadena ambigua que permitía dos esquemas de segmentación: (a) una segmentación en la que los elementos frecuentes aparecían en posición final (*XbYcZaXbYc…*) reflejando el orden relativo de los elementos frecuentes (elementos funcionales) e infrecuentes (elementos de contenido) en las lenguas OV, (b) una segmentación en la que los elementos frecuentes aparecían en posición inicial (*aXbYcZaXbY…*) reflejando el orden relativo de elementos frecuentes e infrecuentes en las lenguas VO.

Como en Gervain (2007), el experimento consistía en una familiarización de 17 minutos y 30 segundos, tras la cual los participantes escuchaban pares de estímulos hexasilábicos. Cada par contaba con un estímulo en el cual un elemento frecuente ocupaba la posición final (*XbYcZa*) y un estímulo en el cual un elemento frecuente ocupaba la posición inicial (*aXbYcZ*). Los participantes debían elegir qué estímulo de cada par consideraban que pertenecía a la lengua que habían escuchado durante la familiarización. A diferencia de la lengua artificial empleada en Gervain (2007), sintetizada con una voz en castellano (es1, MBROLA), la lengua artificial en el experimento A se sintetizó con una voz en alemán (de6, MBROLA), lengua desconocida para todos los participantes.

En este experimento se examinaron las preferencias de segmentación de hablantes monolingües de francés, inglés y castellano y hablantes bilingües L1euskera/L2castellano y L1castellano/L2euskera. Cada grupo de hablantes bilingües se subdividió a su vez en dos subgrupos dependiendo de la lengua del contexto, a saber, la lengua en la que los participantes recibían las instrucciones del experimento (castellano, euskera).

En el experimento B, presentado en el capítulo 4, se creó una nueva lengua artificial que compartía con el experimento A la estructura de la lengua y de los estímulos experimentales, así como el procedimiento. No obstante, esta nueva lengua estaba sintetizada con la voz del castellano empleada en Gervain (2007), en lugar de la voz del alemán empleada en el experimento A. Asimismo, la duración de la familiarización en la lengua empleada en el experimento B se redujo de 17 minutos y 30 segundos a 9 minutos y 3 segundos. Un experimento de control (capítulo 4, sección 4.4) descartó una potencial influencia de esta reducción en la duración de la familiarización en la preferencia de segmentación de los participantes. Por último, se creó un nuevo léxico de sílabas CV en el que se evitaron potenciales efectos creados por las restricciones fonotácticas y la variación alofónica que caracterizan las lenguas examinadas. Un análisis de la frecuencia silábica de las sílabas CV empleadas en los experimentos A y B (capítulo 4, sección 4.5.1) descartó una potencial influencia de este cambio en el inventario de sílabas en la preferencia de segmentación de los participantes.

En este experimento se examinaron las preferencias de segmentación de hablantes monolingües de castellano y hablantes bilingües euskera/castellano, subdivididos en cuatro subgrupos a semejanza de los subgrupos examinados en el experimento A.

En el experimento C, presentado en el capítulo 5, se empleó la lengua artificial diseñada en el experimento B, y se introdujo una pista prosódica consistente en una subida de la frecuencia fundamental de 20 Hz en las sílabas infrecuentes, pista que, según la propuesta de Nespor et al. (2008) es característica de las lenguas OV y que, según Bion et al. (2011), debería conllevar una segmentación en la que los elementos prominentes (las sílabas infrecuentes) aparecen en posición inicial, resultando por tanto en una segmentación de la lengua artificial en la que los elementos frecuentes ocupan una posición final. Esta segmentación señalada por la pista prosódica coincidía así con la preferencia de segmentación señalada por la pista estadística a los hablantes de lenguas OV. En cambio, la segmentación señalada por la pista prosódica se encontraba en conflicto con la segmentación señalada por la pista basada en la frecuencia a los hablantes de lenguas VO, a saber, una segmentación en la que los elementos frecuentes ocupaban la posición inicial.

En este experimento se examinaron las preferencias de segmentación de un grupo de bilingües L1euskera/L2castellano, el cual recibió las instrucciones del experimento en euskera, y un grupo de bilingües L1castellano/L2euskera, el cual recibió las instrucciones en castellano.

Resultados

El experimento A, presentado en el capítulo 3, examinaba la preferencia de segmentación de la lengua artificial empleada en Gervain (2007), aunque sintetizada con una voz del alemán. Esta lengua contenía por tanto una pista basada en la frecuencia: la alternancia de sílabas frecuentes e infrecuentes. En este experimento se obtuvieron tres resultados principales[1]:

(a) La lengua del contexto moduló la preferencia de segmentación de los hablantes bilingües L1euskera/L2castellano, en la dirección predicha por la *frequency-based bootstrapping hypothesis.* Así, los hablantes que recibieron las instrucciones en euskera eligieron una segmentación en la que los elementos frecuentes aparecían en posición final con mayor frecuencia que los hablantes que recibieron las instrucciones en castellano. No obstante, este efecto de contexto no se encontró en la preferencia de segmentación de los dos grupos de bilingües L1castellano/L2euskera.

(b) Los hablantes de la lengua OV (euskera) eligieron una segmentación en la que los elementos frecuentes ocupaban una posición final con mayor frecuencia que los hablantes de las lenguas VO examinadas (francés, inglés, castellano), cumpliendo así las

[1] En esta sección se presentan los resultados obtenidos en los tres experimentos realizados con el fin de que puedan ser consultados independientemente. Las motivaciones para cada experimento serán expuestas en la sección dedicada a la discusión.

predicciones de la *frequency-based bootstrapping hypothesis* propuesta por Gervain (2007). No obstante, los hablantes de lenguas VO no mostraron la preferencia de segmentación en la cual los elementos frecuentes ocupan una posición inicial encontrada en los hablantes de lenguas VO en Gervain (2007), sino una preferencia de segmentación en la cual los elementos frecuentes aparecen en posición final, aunque menor que la observada en los hablantes de la lengua OV (euskera).

(c) Todos los grupos de participantes mostraron un sesgo general hacia una segmentación en la cual los elementos frecuentes ocupaban una posición final de la lengua artificial.

El experimento B, presentado en el capítulo 4, investigaba la preferencia de segmentación de una nueva lengua artificial con la misma estructura que la lengua empleada en el experimento A, aunque sintetizada con la voz en castellano empleada en Gervain (2007) en lugar de con la voz en alemán empleada en el experimento A. Esta lengua contenía por tanto la misma pista basada en la frecuencia presente en el experimento A. En este experimento se obtuvieron tres resultados principales:

(a) No se encontraron diferencias en las preferencias de segmentación entre los hablantes de la lengua OV (euskera) y los hablantes de la lengua VO (castellano). No se replicaron por tanto los resultados obtenidos por Gervain (2007) y en el experimento A, ni se cumplieron las predicciones de la *frequency-based bootstrapping hypothesis*.

(b) Los hablantes L1euskera/L2castellano que recibieron las instrucciones en euskera eligieron una preferencia de segmentación en la que los elementos frecuentes ocupaban una posición final con mayor frecuencia que los hablantes L1euskera/L2castellano que recibieron las instrucciones en castellano, aunque esta tendencia no llegó a alcanzar significación ($p = .091$). Al igual que en el experimento A, la lengua del contexto no moduló la preferencia de segmentación de los hablantes L1castellano/L2euskera.

(c) Todos los grupos de participantes mostraron un sesgo general hacia una segmentación en la que los elementos frecuentes ocupaban una posición inicial de la lengua artificial, preferencia que difería significativamente de la preferencia por una segmentación en la que los elementos frecuentes ocupaban una posición final obtenida en el experimento A.

El experimento C, presentado en el capítulo 5, examinaba la preferencia de segmentación de la lengua artificial empleada en el experimento B, con la manipulación añadida de una subida de la altura tonal de 20 hercios en cada sílaba infrecuente. Esta lengua contenía por tanto la misma pista basada en la frecuencia presente en los experimentos A y B, además de la mencionada pista prosódica. En este experimento se obtuvieron los siguientes resultados:

(a) La presencia de la pista prosódica (cambios en la altura tonal), propuesta como correlato acústico del valor de cabeza final del parámetro de direccionalidad de la cabeza sintáctica no conllevó una mayor preferencia de segmentación en la que los elementos frecuentes ocupaban una posición final (y por tanto con prominencia inicial) en ninguno de los dos grupos de bilingües (L1euskera/L2castellano y L1castellano/L2euskera). Ambos grupos de bilingües mostraron una preferencia de segmentación similar a la obtenida en el experimento B.

(c) No se encontraron diferencias en la preferencia de segmentación entre los dos grupos de bilingües. Así, la información convergente proporcionada por las pistas estadística y prosódica a los hablantes de lenguas OV no conllevó una mayor preferencia de segmentación en la que los elementos frecuentes ocupaban una posición final.

Discusión general

Como se mencionó en la primera sección del presente resumen, tres eran los objetivos principales de esta tesis doctoral: (a) examinar si los hablantes bilingües adultos implementan las estrategias de segmentación prosódicas y estadísticas que caracterizan sus dos lenguas, (b) explorar el papel que juegan las pistas estadísticas y prosódicas en la segmentación de habla en hablantes adultos, y (c) observar si los hablantes adultos emplean las dos pistas que, según Gervain (2007) y Nespor et al. (2008) podrían permitir a los niños establecer el valor del parámetro de direccionalidad de la cabeza sintáctica, a saber, la distribución de la frecuencia de los elementos funcionales y elementos de contenido en las lenguas naturales (pista basada en la frecuencia) y la prominencia relativa dentro de los sintagmas fonológicos (pista prosódica).

La implementación de las estrategias de segmentación en hablantes bilingües

Los experimentos A, B y C exploraban el primero de estos objetivos. Así, los experimentos A y B se centraban en una pista de segmentación estadística, la frecuencia. Los resultados de los experimentos A y B mostraron que las preferencias de segmentación de la lengua artificial de los hablantes bilingües L1euskera/L2castellano estaban moduladas por la lengua del contexto (la lengua en la que los participantes recibían las instrucciones), en la dirección predicha por la *frequency-based bootstrapping hypothesis*, a diferencia de las preferencias de segmentación de los hablantes bilingües L1castellano/L2euskera. Estos resultados sugieren que los bilingües L1euskera, a diferencia de los bilingües L1castellano, fueron capaces de emplear las estrategias de segmentación de su L1 (euskera) y su L2 (castellano).

En el capítulo 3, sección 3.3.1 y capítulo 4, sección 4.5.3, argumenté que esta asimetría encontrada en las preferencias de segmentación entre los hablantes bilingües L1euskera y L1castellano podría provenir de diferencias en la competencia de la L2 entre estos dos grupos de bilingües, como resultado de la situación de diglosia característica del País Vasco, región en la cual el castellano es la lengua preponderante. Estas diferencias en la competencia podrían haber conllevado diferencias en el procesamiento, en concordancia con los resultados obtenidos en estudios sobre acceso léxico, los cuales han mostrado que las diferencias en la competencia conllevan diferencias en el procesamiento.

Alternativamente, sugerí que la estrategia de segmentación en la que los elementos frecuentes ocupan una posición inicial podría ser la estrategia universalmente preferida o no marcada. Si, en consecuencia, la segmentación en la cual los elementos frecuentes ocupan una posición final fuera en efecto la estrategia marcada, podría ser implementada solamente si formara parte del inventario de estrategias que caracteriza la L1, explicando así la asimetría encontrada entre los dos grupos de bilingües. Kayne (1994) propone que todas las lenguas comparten un orden de palabras universal subyacente en el cual las cabezas preceden a sus complementos. La segmentación en la que los elementos frecuentes ocupan una posición inicial es la estrategia que se correlaciona con un orden de cabezas y complementos en el que la cabeza precede al complemento. Los hablantes bilingües L1castellano (lengua de cabeza inicial), por tanto, serían incapaces de implementar la estrategia marcada, característica de su L2 (euskera, lengua de cabeza final).

El experimento C examinaba la interacción de las pistas prosódicas y estadísticas en la segmentación de una lengua artificial por parte de hablantes bilingües euskera/castellano. La lengua artificial contenía cambios en la altura tonal (una subida tonal en cada sílaba infrecuente). Según Nespor et al. (2008), estos cambios en la altura tonal deberían conllevar una segmentación en la que la prominencia se encuentra en posición inicial, resultando en una segmentación en la que los elementos frecuentes ocupaban una posición final. Esta pista se ha propuesto como el correlato acústico de las lenguas de cabeza final. Por tanto, se predecía una segmentación en la que los elementos frecuentes ocupaban una posición final en hablantes L1euskera/L2castellano, es decir, en hablantes nativos de una lengua de cabeza final. Una segmentación de la lengua en la que los elementos frecuentes ocupan una posición final en los hablantes bilingües L1castellano/L2euskera implicaría que estos hablantes bilingües habrían implementado el procedimiento de segmentación de su L2.

Los resultados del experimento C mostraron que la inserción de esta pista prosódica no conllevó una preferencia de segmentación diferente a la obtenida en la lengua artificial prosódicamente plana empleada en el Experimento B en ninguno de los dos grupos de hablantes bilingües. Dada la aparente insensibilidad hacia la pista prosódica obtenida en ambos grupos, es imposible extraer ninguna conclusión sobre las habilidades potenciales de los bilingües L1castellano/L2euskera para emplear la estrategia de segmentación que caracteriza su L2.

Las pistas estadísticas y prosódicas en la jerarquía de pistas de segmentación

El segundo objetivo de los experimentos presentados era investigar el peso relativo que los hablantes adultos otorgan a las pistas estadísticas y prosódicas en comparación con otras pistas de segmentación.

El experimento A exploró la sensibilidad de los hablantes adultos a las pistas estadísticas. En este experimento se obtuvo una preferencia general en todos los grupos de participantes por una segmentación en la que los elementos frecuentes ocupaban una posición final. Argumenté que las propiedades acústico-fonéticas del alemán, lengua empleada para sintetizar los estímulos, podría haber causado este sesgo: los valores del VOT (Voice Onset Time) de las oclusivas sordas podrían haber llevado a los participantes a segmentar la cadena en el inicio de estas oclusivas, las cuales aparecían siempre en el ataque silábico de las sílabas infrecuentes, resultando así en una segmentación en la que los elementos frecuentes ocupaban una posición final de la lengua artificial.

La preferencia general por una segmentación en la que los elementos frecuentes ocupaban una posición inicial obtenida en el experimento B proporcionó evidencia que apoyaba esta propuesta. La lengua artificial en el experimento B tenía una estructura similar a la lengua en el experimento A, pero estaba sintetizada con una voz en castellano, eliminando las pistas acústico-fonéticas que contenía la voz en alemán. Este cambio de una preferencia de segmentación en la que los elementos frecuentes ocupaban una posición final (experimento A) a una preferencia de segmentación en la que los elementos frecuentes ocupaban una posición inicial (experimento B) sugiere que las pistas acústico-fonéticas presentes en el input determinaron la preferencia de segmentación de los participantes, y no las pistas estadísticas, resultado que coincide con los obtenidos por Fernandes, Ventura y Kolinsky (2007) y que apoya una concepción jerárquica de las pistas de segmentación tal y como proponen Mattys et al. (2005).

No obstante, las diferencias en la preferencia de segmentación de la lengua artificial obtenidas entre los hablantes de lenguas VO y OV, las cuales mostraban la dirección predicha por la *frequency-based bootstrapping hypothesis*, revelan que las pistas estadísticas modularon la preferencia de segmentación de los hablantes adultos.

La lengua artificial en el experimento C contenía una pista prosódica, además de la pista basada en la frecuencia presente en los experimentos A y B. Estas pistas proporcionaban información en conflicto a los hablantes L1castellano/L2euskera, e información convergente a los hablantes L1euskera/L2castellano. Una diferencia en la preferencia de segmentación entre estos dos grupos habría revelado por tanto que los participantes habrían integrado estas dos pistas. Los resultados del experimento C mostraron que la información prosódica no influyó en la preferencia de segmentación de ninguno de los grupos y por tanto no se obtuvo dicha diferencia, lo cual sugiere que los participantes no se apoyaron en las pistas prosódicas para segmentar el input.

Esta insensibilidad hacia la pista prosódica podría resultar de potenciales pistas acústico-fonéticas presentes en el input. Dado que las lenguas artificiales en los experimentos B y C contenían la misma información acústico-fonética, el resultado obtenido sugiere que las pistas prosódicas se habrían visto sobrepasadas por la información segmental, tal y como predicen Mattys et al (2005). No obstante, es imposible determinar si (a) los participantes integraron la información de estas dos pistas pero, al pertenecer las pistas estadísticas y prosódicas a niveles inferiores en la jerarquía de pistas de segmentación que la información segmental, no bastaron para sobrepasar la información segmental presente en el input; o si (b) los participantes no fueron capaces de integrar estas dos pistas.

Alternativamente, si, tal y como propuse en la sección anterior, la segmentación en la cual los elementos frecuentes ocupan una posición inicial es la estrategia de segmentación universalmente preferida, el resultado del experimento B sugeriría que ni la información prosódica ni la información convergente proveniente de pistas prosódicas y estadísticas podrían sobreponerse a esta preferencia de segmentación por defecto. En cualquier caso, los resultados del experimento C revelan que la información prosódica parece jugar un papel menor en la segmentación de habla por hablantes adultos.

El papel de las pistas estadísticas y prosódicas en el establecimiento del valor del parámetro de direccionalidad de la cabeza sintáctica

El tercer objetivo de la presente tesis doctoral era examinar si los hablantes adultos pueden emplear dos tipos de información que, como se ha propuesto, podrían permitir a los niños descubrir el orden de palabras básico de la lengua que están adquiriendo y establecer el valor del parámetro de direccionalidad de la cabeza sintáctica: la distribución de la frecuencia de los elementos funcionales y los elementos de contenido en las lenguas naturales (pista basada en la frecuencia) y la prominencia relativa dentro de los sintagmas fonológicos (pista prosódica).

El experimento A investigaba si los hablantes de lenguas VO y OV empleaban la pista estadística proporcionada por la distribución de la frecuencia de los elementos frecuentes e infrecuentes, tal y como predice la *frequency-based bootstrapping hypothesis* propuesta por Gervain (2007), y mostraban por tanto diferentes preferencias de segmentación. Los resultados del experimento A revelaron que los hablantes de la lengua OV examinada (euskera) mostraron una preferencia de segmentación en la que los elementos frecuentes ocupaban una posición final con mayor frecuencia que los hablantes de lenguas VO (francés, inglés, castellano), cumpliendo así las predicciones de esta hipótesis. Este resultado muestra que los hablantes adultos son capaces de computar la frecuencia relativa de los elementos de la lengua artificial y hacer uso de elementos similares a los funcionales en términos de frecuencia para dividir el input en unidades similares a sintagmas.

No obstante, los hablantes de lenguas VO no mostraron la esperada preferencia de segmentación en la que los elementos frecuentes ocupaban una posición inicial, sino una segmentación en la que los elementos frecuentes ocupaban una posición final aunque menor que la de los hablantes de la lengua OV, probablemente a causa de la información acústico-fonética presente en la lengua artificial.

En el experimento B no se encontró la diferencia en la preferencia de segmentación predicha por Gervain (2007) y obtenida entre los hablantes de lenguas VO (castellano) y OV (euskera) en el experimento A y en Gervain (2007), sino una preferencia de segmentación en la que los elementos frecuentes ocupaban una posición inicial en ambas poblaciones. Esta preferencia de segmentación obtenida en el grupo de hablantes nativos de euskera (OV) es sorprendente, dado que la lengua artificial empleada en el experimento B tenía la misma estructura y estaba sintetizada con la misma voz empleada en Gervain (2007) y revela la necesidad de nuevas investigaciones.

El experimento C investigaba una pista prosódica que, según se ha propuesto, podría ayudar a los niños a establecer el valor del parámetro de direccionalidad de la cabeza sintáctica, a saber, la localización y realización de la prominencia principal dentro de los sintagmas fonológicos. Este experimento examinaba si los hablantes bilingües euskera/castellano mostraban sensibilidad hacia la pista prosódica consistente en cambios en la altura tonal. Según Nespor et al. (2008) y Bion et al. (2011), las sílabas con mayor altura tonal se perciben como sílabas iniciales de un constituyente, en un ritmo trocaico, y este ritmo es característico de las lenguas de cabeza final, como el euskera. En estas lenguas la mayor prominencia recae sobre el complemento situado a la izquierda del núcleo. Esta pista conllevaría por tanto una preferencia de segmentación en la cual las sílabas infrecuentes ocuparan una posición inicial (donde una sílaba infrecuente equivaldría en términos de frecuencia a un complemento). Los resultados de este experimento revelaron que ninguno de los dos grupos de hablantes bilingües examinados (L1euskera/L2castellano, L1castellano/L2euskera) eligió la segmentación esperada.

Los experimentos presentados muestran que los hablantes adultos pueden hacer uso de las estrategias de segmentación basadas en la frecuencia, aunque éstas se ven sobrepasadas con facilidad por otras pistas presentes en el input. Asimismo, los resultados del experimento C revelan que los hablantes adultos no hacen uso de la pista prosódica propuesta como potencialmente de ayuda para emprender la adquisición de la sintaxis. La influencia de estas dos pistas parece por tanto ser limitada en la segmentación del habla en los adultos.

Conclusiones

→ Los hablantes bilingües pueden emplear las estrategias de segmentación basadas en la frecuencia características de sus dos lenguas, aunque la adquisición de la estrategia de segmentación de la L2 parece estar limitada. Queda pendiente determinar las causas de esta limitación.

→ Las pistas estadísticas y prosódicas parecen jugar un rol menor en la segmentación del habla de los hablantes adultos. Otras pistas tales como las pistas acústico-fonéticas se superponen a las estadísticas y prosódicas, apoyando así una concepción jerárquica de las pistas de segmentación según la cual la información estadística y la prosódica son las pistas de segmentación a las que los hablantes adultos atribuyen menor peso.

→ La segmentación en la cual los elementos frecuentes ocupan una posición inicial podría ser la estrategia de segmentación universalmente preferida. Futuras investigaciones se ocuparán de examinar esta hipótesis.

→ Las estrategias de segmentación basadas en la frecuencia son pistas disponibles para los hablantes adultos en la segmentación del habla. Así, los hablantes adultos computan la distribución de la frecuencia de los elementos, y el orden relativo de los elementos frecuentes (elementos funcionales) e infrecuentes (palabras de contenido) en sus lenguas nativas modula las preferencias de segmentación de los hablantes.

Chapter 1. Introduction

Words and phrases are not discrete events in speech, that is, word and phrase boundaries are not consistently marked by cues such as pauses (Cole and Jakimik, 1980). Notwithstanding this continuous character of speech, speech segmentation is an automatic and unconscious process for adult listeners and a task accomplished seemingly effortlessly by infants. Infants must locate the boundaries to words and phrases in order to acquire the lexicon and syntax of language. Corpora analysis has revealed that infant-directed speech typically consists of multi-word utterances, and that the words that occur in isolation (e.g., *yeah*, *ok*) do not tend to appear in multiword utterances (Aslin, Woodward, LaMendola and Bever, 1996; Cairns, Shilcock, Chater and Levy, 1994). Moreover, even when explicitly instructed to teach their infants new words, mothers frequently do not present these words in isolation (Woodward and Aslin, 1990, 1991, 1992, as cited in Aslin, Woodward, LaMendola and Bever, 1996).

Speech segmentation results from the acquisition and integration of various perceptual strategies that take place during development (Christiansen, Allen and Seidenberg, 1998; Christiansen, Conway and Curtin, 2005). During the last decades, a substantial amount of research has investigated the cues present in the input that allow humans, both adults and infants, to achieve successful speech segmentation. This research has examined cues to word segmentation and cues to the segmentation of speech into bigger syntactic constituents, such as phrases. The former are crucial for the infants' building of the lexicon, whereas the latter enable infants to divide the input into smaller units and detect regularities present in the input, which might lead infants to bootstrapping the acquisition of syntactic rules. Prosodic, statistical, phonotactic and acoustic-phonetic cues such as coarticulation and allophonic variation have stood out as available cues used by both adults and infants to break speech into words and phrases (Bagou, Fougeron and Frauenfelder, 2002; Cutler and Norris, 1988, Fernandes, Ventura and Kolinsky, 2007; Friederici and Wessels, 1993; Hirsh-Pasek et al., 1987; Johnson and Jusczyk, 2001; Jusczyk, Hohne and Bauman, 1999; Jusczyk, Houston and Newsome, 1999; McQueen, 1998; Mehler, Dommergues, Frauenfelder and Seguí, 1981; Otake, Hatano, Cutler and Mehler, 1983; Saffran, Aslin and Newport, 1996; Saffran, Newport and Aslin, 1996; Smith and Hawkins, 2000).

Successful deployment of these cues often requires knowledge of the specific properties of the native language. Thus, although sensitivity to rhythmic properties of the languages appears to emerge even prior to birth—near-term foetuses are able to discriminate between their native language and a foreign language with different rhythmic properties (Kisilevsky et al., 2009)—, infants only start segmenting words that conform the native language's rhythm at around 7½ months of age (Jusczyk, Houston and Newsome, 1999).

Similarly, phonotactic constraints and allophonic variation, two cues used both by adults and infants in speech segmentation, vary cross-linguistically. It is hence not surprising that the use of these two cues in speech segmentation emerges at around 9 to 10½ months of age (Friederici and Wessels, 1993; Jusczyk, Hohne and Bauman, 1999), a time at which perceptual reorganization is known to take place (Bosch and Sebastián-Gallés, 2003; Polka and Werker, 1994; Werker and Lalonde, 1988; Werker and Tees, 1984). During this perceptual reorganization, a tuning to the native languages' properties takes place in parallel to a decline in sensitivity toward features that do not characterize the native language. At around 10 to 12 months of age infants' ability to discriminate non-native speech sound contrasts declines, as a function of the specific properties of the language under acquisition (Werker and Tees, 1984). This language-specific knowledge, essential to the successful use of the segmentation cues, opens the ground for investigating the bilinguals' ability to deploy the segmentation strategies characteristic of their two languages.

Among the cues that assist in speech segmentation, two appear to be crucial at the onset of language acquisition, namely prosodic and statistical information. It has been proposed that these two cues not only assist infants in the location of word and phrase boundaries, but they also provide information that may allow infants set the values of basic grammatical properties of the language under acquisition, hence bootstrapping the acquisition of syntax (Christophe, Guasti, Nespor, Dupoux and van Ooyen, 1997; Gervain, 2007; Mazuka, 1996; Nespor, Guasti and Christophe, 1996). Throughout development, the relative weight of these cues changes, favouring segmental cues (e.g., acoustic-phonetic and phonotactic cues) and lexical information and relegating statistical and prosodic cues to a less salient position in adult speech segmentation (Fernandes, Ventura and Kolisky, 2007; Mattys, White and Melhorn, 2005).

Three are the main aims of the present dissertation. First, it aims to investigate the abilities of adult bilingual speakers to deploy the prosodic and statistical segmentation strategies that characterize their two languages. The present dissertation aims thus to contribute to the existing literature on bilingual language processing. Additionally, in this dissertation I aim to further explore the role that statistical and prosodic cues play in adult speech segmentation, within the framework of a hierarchical account of segmentation cues. Last, this dissertation aims to examine whether adult speakers can make use of two types of information that have been proposed as potentially allowing infants to figure out the basic word order of the language under acquisition and set the value of the Head-Directionality parameter (Baker, 2001; Chomsky, 1981): the frequency distribution of functors and content words in natural languages (i.e., a frequency-based cue) and the relative prominence within phonological phrases (i.e., a prosodic cue) (Christophe et al., 1997; Gervain, 2007; Nespor, Guasti and Christophe, 1996; Nespor et al., 2008).

To that end, the results of three artificial language learning experiments are presented, in which I examine the segmentation preferences of ambiguous artificial languages that contain these frequency-based and prosodic cues, by adult monolingual and bilingual speakers.

This dissertation is organized as follows. In Chapter 2, I review the literature on infants and adults' abilities to make use of the statistical and prosodic information contained in the input. A first section in Chapter 2 concentrates in adults and infants' statistical learning abilities and pays special attention to the statistical and distributional properties of functors in natural languages. A subsequent section focuses on the sensitivity to different types of prosodic information displayed by adults and infants. A last section briefly describes other cues used in speech segmentation and presents the hierarchical account of segmentation cues proposed by Mattys, White and Melhorn (2005). According to their proposal, segmentation cues are hierarchically organized depending on the relative weight that speakers give to each cue, and this relative weight changes through linguistic and cognitive development.

In Chapter 3, I present and discuss an artificial language learning experiment which examines whether adult bilingual speakers are able to implement the frequency-based segmentation procedures that characterize their two languages. In this experiment I show that the language of the context (that is, the language in which participants are given the instructions) modulates the segmentation preference of bilingual speakers. Also, the results of this experiment reveal that adult speakers track the frequency distribution of frequent and infrequent elements in the input and that speakers of head-initial (VO) and head-final (OV) languages display different segmentation preferences of an ambiguous language, as predicted by the frequency-based bootstrapping hypothesis (Gervain, 2007).

In Chapter 4, I present and discuss the results of a second artificial language learning experiment, which explores the interplay of statistical and acoustic-phonetic cues in adult speech segmentation. The results of this experiment show that adult speakers weigh acoustic-phonetic cues more heavily than statistical cues, a result that supports a hierarchical account of segmentation cues, as proposed by Mattys et al. (2005). Also, this experiment obtains the language context effect observed in the experiment presented in Chapter 3.

In Chapter 5, I present and discuss the results of a third artificial language learning experiment, which investigates the interplay between statistical and prosodic cues in adult speech segmentation. In this experiment I show that prosodic cues appear to play a minor role in adult speech segmentation. Specifically, the results of this experiment reveal that adult speakers do not make use of a type of prosodic information in speech segmentation which has been proposed as a potential bootstrapping mechanism to the acquisition of the basic word order of the language during language acquisition: the relative prominence within phonological phrases.

Last, in Chapter 6, I discuss the findings obtained in the experiments presented in Chapters 3, 4 and 5, addressing separately the three topics presented in the aims of the present dissertation: bilinguals' processing of the segmentation strategies of the L2, the status of statistical and prosodic cues in the hierarchy of segmentation cues, and the availability of frequency-based and prosodic cues to the basic word order in adult speech segmentation.

Chapter 2. Chunking the input

The present chapter discusses the ample research conducted in the last decades regarding adults' and infants' sensitivity to two segmentation cues, namely, statistical and prosodic information. This research shows that sensitivity to statistical and prosodic cues emerges very early in development, and appears to play a crucial role in the earliest stages of language acquisition. Prosodic and statistical information are the initial cues used by infants to locate word boundaries in the input, helping infants start building the lexicon. Furthermore, these two cues assist infants and adults in the segmentation of speech into bigger syntactic constituents, such as phrases, providing speakers and learners with a representation of the structure of the input and allowing them to discover the regularities present in it. Prosodic cues and statistical and distributional properties of functional elements appear to be of special importance in bootstrapping the acquisition of syntax, and might allow infants to set the values of major syntactic parameters prior to lexical knowledge.

This chapter focuses additionally on the interplay between statistical and prosodic cues and other segmentation cues available to adults and infants throughout development (i.e., phonotactic constraints, coarticulation, allophonic variation). Studies on the relative weight given to these cues in speech segmentation have revealed that they appear to be hierarchically organized, as proposed by Mattys, White and Melhorn (2005). Adults and infants rely on different cues to accomplish the task of speech segmentation. While adult speakers primarily rely on lexical and segmental cues (e.g, phonotactic constraints, coarticulation) rather than on prosodic and statistical cues, infants weigh more heavily those cues that do not require fine-grained knowledge of the properties of the language under acquisition (i.e., prosodic and statistical cues).

2.1 Cues related to frequency

2.1.1 Statistical learning abilities in adults and infants

Statistical learning is a mechanism used by both infants and adults to extract the regularities present in the input. Statistical learning is a domain-general capacity not specific to language, i.e., listeners use statistical learning to discover regularities in non-linguistic auditory stimuli, e.g., for segmentation of a tone stream (Saffran, Johnson, Aslin and Newport, 1999) and in other modalities like with visual stimuli (Fiser and Aslin, 2001; Kirkham, Slemmer and Johnson, 2002). Furthermore, the ability to track distributional regularities is not human-specific: non-human mammals like cotton-top tamarins and rats are also able to extract distributional regularities from speech streams, albeit this ability is more limited than in humans (Hauser, Newport and Aslin, 2001; Toro and Trobalón, 2005).

A type of statistical learning picked up both by adults and infants is the distribution of phonemes and sequences of phonemes in the native language, essential to the acquisition of phonotactic constraints, the language-specific restrictions on the occurrence of phonemes and sequences of phonemes both within and across words. Adults and infants make use of the native language's phonotactic constraints as a cue to the location of word boundaries, since sequences that are phonotactically illegal within a word cue the presence of a word boundary (Friederici and Wessels, 1993; McQueen, 1998). In addition to tracking the legality of phonemes or sequences of phonemes, infants and adults make use of probabilistic phonotactics in word segmentation, i.e., listeners compute the relative frequency of occurrence of legal phoneme sequences (Mattys and Jusczyk, 2001; van der Lugt, 2001). van der Lugt (2001) showed that Dutch adult speakers detected target Dutch monosyllables embedded in nonsense sequences faster when the targets started with a CV sequence with high frequency of occurrence after a word boundary than when the targets started with a CV sequence with low frequency of occurrence after a word boundary.

By 9 months of age, infants have acquired fine-grained knowledge of the distribution of phonemes and sequences of phonemes in the language under acquisition (Friederici and Wessels, 1993; Jusczyk, Luce and Charles-Luce, 1994). Thus, 9-month-olds prefer to listen to lists of nonwords that conform to the phonotactic constraints in the native language rather than to lists of nonwords that contain illegal phonotactic sequences (Friederici and Wessels, 1993; Jusczyk, Friederici, Wessels, Svenkerud and Jusczyk, 1993). Furthermore, Mattys and Jusczyk (2001) showed that 9-month-old infants segmented CVC targets from passages when the targets occurred surrounded by high-frequency word-boundary phonotactic cues better than when the targets were surrounded by low-frequency word-boundary phonotactic cues.

Chambers, Onishi and Fisher (2003) showed that infants can learn new phonotactic regularities after brief exposure. They familiarized 16½-month-old infants during 3 to 4 minutes with lists of CVC syllables that contained two sets of consonants, each set restricted to either the onset or the coda position. After familiarization, infants listened to previously unheard CVC syllables that either followed or violated the new phonotactic constraints. The infants listened longer to the illegal items than to the legal items, i.e., infants successfully learnt the phonotactic restrictions and generalized these to new syllables. The statistical learning necessary to the acquisition of phonotactic constraints is not exclusively used during acquisition, but is also available to adults. Thus, Onishi, Chambers and Fisher (2002) showed that, as infants, adult participants were able to learn and generalize new phonotactic constraints after only brief exposure to monosyllabic tokens that contained the new phonotactic patterns.

Transitional Probabilities (TPs) between adjacent syllables are another type of statistical learning which has been proposed as an available cue to word segmentation in both infants and adults (Saffran, Aslin and Newport, 1996; Saffran, Newport and Aslin, 1996). Two are the computations necessary to track TPs: (a) the relative frequency of a given segment or sequence of segments (X) (e.g., a syllable) in the speech stream and (b) the frequency of co-occurrence of this particular segment or sequence of segments and another segment or sequence of segments (XY). The rationale behind TPs is that the probability that syllable X is followed by syllable Y is higher when these two syllables occur word-internally than when they span a word boundary. Thus, in the sequence *pretty baby*, TPs between the syllables *pre* and *ty* and between *ba* and *by* will be higher than between *ty#ba* (Saffran, Aslin and Newport, 1996). Saffran, Newport and Aslin (1996) formalize the computation of the TP of a syllable *Y* given a syllable *X* as follows:

Figure 1. Computation of TPs. From Saffran et al. (1996: 1928)

$$\text{Y|X} = \frac{\text{frequency of pair } XY}{\text{frequency of } X}$$

Adults and 5½- to 8-month-old infants segment artificial languages that contain no other information but TPs between adjacent syllables (Johnson and Tyler, 2010; Saffran, Aslin and Newport, 1996; Saffran, Newport and Aslin, 1996). In their seminal study, Saffran, Newport and Aslin (1996) presented adult listeners with an artificial language that consisted of 12 CV syllables (e.g., *ba, pu*), combined into 6 trisyllabic word-like items (e.g., *babupu, bupada*). Participants listened to three 7 minute long strings that consisted of random concatenation of the trisyllables. TPs across item boundaries were lower than TPs within items. After familiarization, participants were presented with pairs of trisyllabic strings and were asked to choose which member of each pair sounded more like the artificial language they heard previously. Each pair contained one of the word-like items, accompanied by either (a) a non-item, i.e., three of the language's syllables presented in a previously unheard order, hence having a TP of 0, or (b) a part-item, i.e., two syllables of an item combined with another syllable from the artificial language. Participants chose the word-like items significantly more often than the non-items and part-items (76% and 65% respectively), i.e., participants segmented the statistically defined chunks of syllables.

Weiss, Gerfen and Mitchell (2009) found that adults exposed to a mixed input of two artificial languages that aimed to emulate language acquisition in a bilingual environment were able to separately compute the TPs of syllables in the two languages. Weiss et al. (2009) presented adult participants with interleaved 2 minute strings of two artificial languages, during a total of 24 minutes. In one condition, TPs between syllables and phonemes remained constant regardless of whether the TPs were computed for each separate language or for the combination of both languages (congruent condition). In another condition, while the computation of TPs for each separate language was constant, the statistical information that resulted from the combination of the two languages yielded less constant TPs, potentially obscuring segmentation of the languages into chunks (incongruent condition). After familiarization, participants listened to pairs of items which consisted of word-like items and part-items (the members of each pair belonged to one of the two languages). Participants in both the congruent and incongruent conditions chose the word-like items more often than the part-items, i.e., they were able to learn the two languages. The fact that participants successfully learnt the word-like items in the incongruent condition suggests that the learners did not unify the statistical information from both languages into a sole representation, but tracked the statistical information of both languages separately, forming two representations.

Saffran, Aslin and Newport (1996) used a simplified version of Saffran, Newport and Aslin's (1996) artificial language to study—using the Headturn-Preference Procedure (HPP)—whether 8-month-old infants were able to segment the input based on TPs between adjacent syllables. Stimuli were four trisyllabic words (e.g., *bidaku*, *padoti*) concatenated in a monotonous 2 minute long stream. Again, TPs in adjacent syllables within word-like items were higher (TPs: 1.0, e.g., *bi.da*) than TPs between word-like items (TPs: 0.33, e.g., *ku#pa*). The 8-month-olds listened longer to the non-items and part-items than to the word-like items, i.e., they showed a novelty effect which revealed that infants had computed the statistical information present in the artificial language.

Segmentation based on the computation of TPs between adjacent syllables has been replicated in numerous studies, both with synthesized and natural stimuli and found with infants as young as 5½ months of age (Aslin, Saffran and Newport, 1998; Johnson and Jusczyk, 2001; Johnson and Tyler, 2010; Jusczyk, Hohne and Bauman, 1999; Jusczyk, Houston and Newsome, 1999; Pelucchi, Hay and Saffran, 2009; Thiessen and Saffran, 2003).

Graf-Estes, Evans, Alibali and Saffran (2007) and Saffran (2001) investigated the nature of infants' representations of the chunks segmented from artificial languages and found that infants appeared to treat the segmented sequences as actual candidates for words. Thus, in two experiments, Graf-Estes et al. (2007) familiarized 17-month-old infants in an artificial language and subsequently tested infants with object-labellearning tasks. Word-like items segmented from the artificial language facilitated infants' learning of new object labels, whereas no such facilitation was found for non-items and part-items. The facilitation found in the segmented word-like items suggests that infants represent these chunks as potential words which can be linked to meanings.

Adults and infants are also able to perform statistical learning of non-adjacent dependencies. These dependencies are essential to the acquisition of structural information, such as morphosyntatic rules in natural languages (e.g., the relation between an auxiliary and an inflectional morpheme on the main verb, as in: *She **is** danc**ing***). Peña, Bonatti, Nespor and Mehler (2002) showed that adult listeners were able to segment an artificial language based on TPs between non-adjacent syllables. Thus, participants were familiarized to a stream that consisted of concatenation of trisyllabic items with the structure A_iXC_i, where the occurrence of A predicted the occurrence of C (as signalled by the subscript $_i$, e.g., ***pu**li**ki**, **pu**fo**ki***) and had hence a TP of 1.0. After familiarization, participants listened to pairs of items that contained word-like items (A_iXC_i) and part-items (C_kA_iX or XC_iA_j) and were asked to choose which items sounded more likely to belong to the language previously heard. Participants chose the word-like items more often than the part-items, that is, they successfully learnt the word-like items. However, TPs did not suffice to discover the rule that underlied the distribution of the non-adjacent syllables and hence generalize this rule to new stimuli. Peña, Bonatti, Nespor and Mehler (2002) familiarized participants to the artificial language described above and subsequently presented them with pairs of items that consisted of a part-item and a previously unheard sequence that followed the A_iXC_i generalization (rule-item). Participants did not choose the rule-items more often than the part-items; they failed to generalize the rule to new but structurally similar stimuli.

In order for adults to learn the rules signalled by non-adjacent dependencies, the input must contain other cues such as pauses. Peña et al. (2002) showed that insertion of 25 millisecond subliminal pauses between the trisyllabic items in the stream allowed participants to discover the rule and generalize it to new stimuli. Furthermore, in a series of artificial language learning experiments, Endress and Mehler (2009) found that in order for adult listeners to learn the generalizations present in the input, the syllables carrying the non-adjacent dependency had to occur at edge positions. Thus, when participants were familiarized with streams containing 1 second pauses between word-like units, they learnt the rule regarding the occurrence of A and C in a language with the structure A_iXYZC_i, but did not learn the same rule in a language with the structure XA_iYC_iZ, that is, when the syllables carrying the non-adjacent dependency (i.e., A_i, C_i) occurred in medial position.

Toro, Nespor, Mehler and Bonatti (2008) found further constraints to the generalization of rules. In a series of artificial language learning experiments, Toro et al. (2008) showed that rules signalled by non-adjacent dependencies were only extracted when vowels carried the dependency, but did not take place when consonants carried this dependency. Toro, Shukla, Nespor and Endress (2008) found that this asymmetry did not originate in acoustical features of vowels and consonants, i.e., adult speakers failed to generalize the rule when consonants carried the dependency even when vowels were perceptually diminished through reduced duration and consonants were made salient through increased duration. Adults could generalize the rule only in the absence of vowels, but the generalization remained marginal.

On the other hand, Bonatti, Peña, Nespor and Mehler (2005) and Toro, Nespor, Mehler and Bonatti (2008) found that listeners relied on consonants for the segmentation of an artificial language, i.e., listeners segmented the word-like items based on TPs between non-adjacent syllables when the dependency was carried by consonants. Pons and Toro (2010) provided further supporting evidence on the different role of consonants and vowels in speech processing. Thus, Pons and Toro (2010) showed that 11-month-old infants extracted simple rules from CVCVCV trisyllabic strings with the structure *AAB* when vowels carried the dependency, but did not generalize the rule when consonants carried this dependency. This result suggests that the different role of consonants and vowels is not restricted to the computation of non-adjacent dependencies.

The constraints on the learnability of non-adjacent dependencies described above show that statistical learning of these dependencies has greater complexity than statistical learning of adjacent dependencies. This complexity is reflected in infants' acquisition of these two types of dependencies. Thus, though 5½- to 8-month-old infants segment speech streams based on adjacent dependencies (Johnson and Tyler, 2010; Saffran, Aslin and Newport, 1996), the ability to compute non-adjacent dependencies seems to develop at around 15 months of age (Gómez and Maye, 2005).

In a series of experiments using the Headturn Preference Procedure (HPP), Gómez and Maye (2005) showed that 15- and 17-month-old infants, but not 12-month-old infants, tracked the non-adjacent dependencies in strings of three word-like items. Thus, infants listened to a 3 minute familiarization to strings of word-like items with the structures *aXc* or *bXd*, i.e., in which the first word-like item of the string (i.e., *a, b*) predicted the presence of the third element (i.e., *c, d*) regardless of the middle element (i.e., *X*). Word-like items within each string (i.e., *a, X* and *c* or *b, X* and *d*) were separated by 250 msec pauses, whereas 750 msec pauses were inserted between the different strings. After exposure, infants listened to trials that (a) had the same structure of the strings in the familiarized artificial language (e.g., *aXc*: *pel kicey rud*, *bXd*: *vot puser jic*), or (b) violated this structure (e.g., *aXd*: *pel kicey jic*, *bXc*: *vot puser rud*). 12-month-old infants did not discriminate between the two types of strings, whereas the groups of 15- and 17-month-old infants discriminated between the two types of strings.

These results suggest that the ability to track non-adjacent dependencies develops between 12 and 15 months of age. However, in order for infants around that age to track non-adjacent dependencies, dependencies in adjacent elements must be highly variable and yield hence low predictability. Gómez (2002) and Gómez and Maye (2005) showed that 17- and 18-month-old infants familiarized to *aXc* and *bXd* strings relied on TPs between adjacent syllables and did not learn the non-adjacent dependency between the first and final syllables when the medial element *X* consisted of a small set of items (3 or 12 items). On the other hand, the infants extracted the non-adjacent dependency and did not rely on the TPs between adjacent syllables when the medial element *X* consisted of a bigger set of items (18 or 24 items), which entailed higher variability of TPs between adjacent syllables.

The evidence summarized so far shows that adults and infants are sensitive to various types of statistical information contained in the input, which enables them to segment speech and discover regularities. This information might allow infants to start building a lexicon and acquiring phonotactic constraints and syntactic regularities that characterize the language under acquisition.

2.1.2 The properties of functional elements

Functional elements signal grammatical relations whereas lexical or content words typically carry lexical meaning. The former form small, closed classes, whereas the latter form open, constantly growing classes with numerous members. Functional elements are characterized by statistical, distributional and phonological properties that set them apart from content words (Morgan, Shi and Allopenna, 1996). It has been proposed that functional elements might play an important role during language acquisition, allowing infants to segment the input into words and phrases (Christophe et al., 1997; Morgan, Meier and Newport, 1987; Shi and LePage, 2008) and discover syntactic regularities (Gervain, 2007; Green, 1979, Morgan et al., 1987; Valian and Coulson, 1988).

The specific phonological and acoustic-phonetic properties of functors vary cross-linguistically and are hence language-specific. However, perceptual minimality has been proposed as a universal characteristic of functional elements. In English, functors tend to have simple onsets and codas (if at all), non-diphthonguized nuclei and may become cliticized (Morgan, Shi and Allopenna, 1996; Selkirk, 1996). Also, English functors are typically unstressed monosyllables that contain reduced vowels, whereas content words usually have an initial stressed syllable realized with a full vowel (Cutler, 1993; Selkirk, 1996). This perceptual minimality attributed to functional elements as opposed to content words has also been attested in cross-linguistic studies on infant-directed speech (Morgan, Shi and Allopenna, 1996, and Shi, Morgan and Allopenna, 1998, for studies on English, Mandarin Chinese and Turkish). Correlation of these phonological cues has been proposed to reliably distinguish functional from content words (Morgan, Shi and Allopenna, 1996). Indeed, studies on word boundary misplacement have shown that adult listeners ascribe different phonological characteristics to functors and lexical words (Cutler and Butterfield, 1992).

Statistically, functors have an extremely high frequency of occurrence. As shown in numerous corpus analyses, functors are characterized by a low type count (i.e., a small number of members), but a very high token frequency (i.e., a very high frequency of occurrence of each member). Content words are characterized by the opposite pattern, a high type count (i.e., a great number of members) and a low token count (i.e., low frequency of appearance of each content word) (Cutler and Carter, 1987;

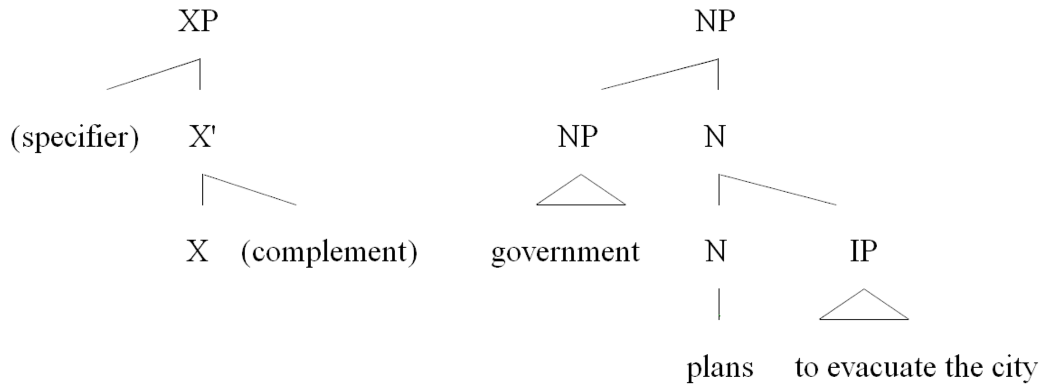
Gervain, 2007; Kučera and Francis, 1967, as cited in Morgan, Shi and Allopenna, 1996; Shi, Morgan and Allopenna, 1998).

Cutler and Carter (1987) showed that functional elements constituted 59% of all word tokens in an English corpus analyzed. However, the English vocabulary contains only a few hundred functors, i.e., less than 1% of the total number of words in the dictionary (Cutler, 1993). Similarly, Gervain (2007) and Gervain et al. (submitted) examined the frequency distributions of functors and content words in corpora of infant-directed speech in Japanese, Basque[2], French and Italian. The 12 to 22 words with highest frequency exclusively corresponded to functors in all languages. These words represented around one third of all word tokens in the corpora (ranging between 29 to 45%).

Distributionally, functors are characterized by a tendency to occur at the edges of syntactic units, such as phrases (Gervain, 2007; Kimball, 1973; Morgan, Shi and Allopenna, 1996). Phrases occupy an intermediate position in the hierarchy of syntactic constituents in natural languages, between words and clauses. Every phrase (e.g., XP) contains a head (X), which determines the syntactic type of the phrase (i.e., the category) and can be accompanied by a complement and/or a specifier. Complements are sisters of heads (X), whereas specifiers are sisters of intermediate projections (X') and daughters of maximal projections (XP). Both content words (e.g., nouns, verbs, adjectives) and functors (e.g., prepositions, auxiliaries) can be heads of syntactic phrases. Thus, in the phrase *government plans to evacuate the city*, the NP *government* occupies the specifier position of the head of the NP, i.e., *plans*, whereas the IP *to evacuate the city* occupies the complement position, as shown in Figure 2.

[2] It should be pointed out that for Basque the corpus analysis was conducted on an extract of the Basque translation of *Harry Potter and the Philosopher's Stone*.

Figure 2. Phrase structure in X-bar theory (where X can represent any category)



The linear order of specifier and head, and complement and head are language-specific, as described by the Head-Directionality parameter (Baker, 2001; Chomsky, 1981). Thus, in head-initial languages (e.g., Spanish, English) heads occur phrase-initially, whereas in head-final languages (e.g., Basque, Japanese), heads occur phrase-finally. Typological studies have shown that the relative order of functors and content words correlates with the order of heads and complements (Greenberg, 1963). In head-initial languages functors tend to appear phrase-initially (e.g., head-initial languages typically have prepositions), whereas in head-final languages functors tend to appear phrase-finally (e.g., head-final languages usually have postpositions).

Though children's early productions typically omit functional elements and have hence been designated as "telegraphic" (Bloom, 1970, as cited in Guasti, 2002; Brown, 1973, as cited in Guasti, 2002), a great deal of evidence has shown a mismatch in the perception and production of functors. The salient properties of functional elements allow infants to sort the words encountered in the input into the functional and lexical categories from an early age. Shi, Werker and Morgan (1999) showed that newborns discriminated between lists of English lexical and functional elements. Not only do telegraphic children have detailed segmental phonological representations of the native language's functors, but this representation seems to emerge gradually between 11 and 13 months of age (Gerken, Landau and Remez 1990; Shafer, Shucard, Shucard and Gerken, 1998; Shi, Werker and Cutler, 2006). 6- to 8-month-old infants segment functional elements, though their representations are not yet fully specified (e.g., infants do not discriminate the functor *the* [*ðə*] from the nonsense functor *kuh* [*kə*]) (Höhle and Weissenborn, 2003; Shi, Cutler, Werker and Cruickshank, 2006; Shi and Gauthier, 2005; Shi, Marquis and Gauthier, 2006).

Infants as young as 8 months of age make use of the knowledge of the native language's functors to segment adjacent content words (Shi, Cutler, Werker and Cruishank, 2006; Shi and LePage, 2008). Shi and Lepage (2008) showed that French-learning 8-month-olds familiarized to (a) a sequence consisting of a French functor followed by a low-frequency noun in infant-directed speech (e.g., *des preuves*) and (b) a sequence consisting of a nonsense functor followed by another low-frequency noun (e.g., *kes sangles*) listened longer to isolated repetitions of the noun that was preceded by the real functor in familiarization than to repetitions of the noun that had been preceded by a nonsense functor.

The statistical properties of functors appear to be crucial to their acquisition. Infants acquire first the functors that occur most frequently in the input and these are represented in detail earlier than less frequent functors (Shi, Cutler, Werker and Cruishank, 2006; Shi, Marquis and Gauthier, 2006). Also, segmentation of adjacent words is initially facilitated only by frequent functors (Shi and LePage, 2008). Furthermore, Hochmann, Endress and Mehler (2010) showed that Italian-learning 17-month-old infants associated an item in a foreign language (French) that had occurred infrequently during familiarization to a content word (a new object) rather than a familiarized item that had occurred more frequently during familiarization. This was an expected result if infants were computing the frequency of occurrence of the items in the familiarization and associating high frequency of occurrence with functional elements and low frequency of occurrence with content words. This preference disappeared in a subsequent experiment that had no familiarization phase. This result suggests that infants were relying on the statistical properties of the items and not on their acoustic properties to categorize new words into the functional and lexical categories.

Additionally, functors have been proposed to help infants acquire certain aspects of syntactic structure. Different functor categories occur adjacent to specific content word categories, i.e., functor categories are often in complementary distribution. Thus, determiners typically occur adjacent to nouns, whereas inflectional morphemes occur adjacent to verbs. It has been proposed that the co-occurrence of specific functional and content word categories might serve as a reliable cue to the categorization of content words into syntactic categories (Gerken and McIntosh, 1993; Golinkoff, Schweisguth and Hirsh-Pasek, 1992, as cited in Golinkoff, Hirsh-Pasek and Schweisguth, 2001). Redington, Chater and Finch (1998) conducted a series of computational experiments with corpora of English infant-directed speech, which showed that the distribution of words in the input contained information that could lead to establishing the syntactic categories. Additionally, Gerken and McIntosh (1993) showed that by 24 months of age infants already had knowledge of the co-occurrence patterns between functional and content word categories. In a picture identification task, infants performed better when the target word (e.g., *bird*) was presented in a sequence that contained functors appropriate to the context (e.g., *find the bird for me*) than when the target was presented in a sequence containing an inappropriate functor (e.g., **find was bird for me*), no functor (e.g., *find bird for me*), or a nonsense syllable (e.g., **find gub bird for me*).

Functors appear to also play an important role in learning syntactic operations of specific languages (Gervain, 2007; Green, 1979, Valian and Coulson, 1988). Green (1979) proposes the Marker Hypothesis, according to which all languages contain a small set of words or morphemes (i.e., functors), each associated with a few syntactic constructions that hence act as markers that signal the occurrence of these specific constructions. The presence of markers to the syntactic structure facilitates parsing the input, whereas its absence hinders processing.

Research on artificial language learning has provided supporting evidence for the Marker Hypothesis (Green, 1979; Morgan, Meier and Newport, 1987). Morgan, Meier and Newport (1987) examined the learnability of artificial languages that contained (a) effective markers to syntactic constituents and rules in the target languages, (b) useless markers, i.e., markers that did not effectively signal syntactic rules or constituents, and (c) no markers. The presence of functor-like markers to phrase-like units facilitated parsing the input and learning the syntactic rules and structure of the target artificial language. Useless markers led to worse performance than presence of effective markers and the absence of markers rendered a significant part of the artificial language's rules virtually unlearnable.

Valian and Coulson (1988) propose that high-frequency markers are used as anchor points that facilitate distributional analysis. To be an effective anchor, a functor must possess two properties: (a) it must be reliably associated with a single syntactic structure, and (b) the marker must have greater frequency than the surrounding items. Valian and Coulson (1988) conducted an artificial language learning experiment and found that participants learnt to distinguish phrasal constituents when the language contained high-frequency markers (i.e., markers that occurred 6 times as often as “lexical-like” items), but not when the language contained low-frequency markers (i.e., markers that occurred 1½ times as often as “lexical” items). Hence, the greater the frequency of occurrence of the markers relative to “lexical” items, the greater the markers’ usefulness as anchor points.

Infants between 15 and 18 months of age start learning the relationships between non-adjacent functors in simple contexts. Santelmann and Juscyzk (1998) found that infants preferred passages that contained well-formed combinations of functors (e.g., *everybody is baking bread*) to passages containing ungrammatical combinations (e.g., **everybody can baking bread*). However, this sensitivity was still restricted to short processing domains. For instance, insertion of long adverbs (e.g., *everybody is cheerfully baking bread*) prevented discrimination.

Gervain (2007) proposed a frequency-based bootstrapping hypothesis, according to which the frequency distribution and relative order of functors and content words in natural languages might allow infants acquire the basic word order of the language under acquisition and, specifically, set one major syntactic parameter, i.e., the Head-Directionality parameter (Baker, 2001; Chomsky, 1981). This bootstrapping hypothesis will be discussed in detail in Chapter 3.

Natural languages contain thus both functional and content words. These two major categories are set apart by an array of phonological, distributional and statistical properties. Functors act as anchoring points that signal syntactic constructions and facilitate regularity extraction, as shown in artificial language learning experiments. The particular properties of functors are salient to the infants from a very early stage in the development. Infants are especially sensitive to the frequency of occurrence of the native language’s functors and acquire first the most frequent functors present in the input. Initial representation of functors is underspecified, building up during the second half of the first year until detailed segmental representation is achieved. Along the second year of life, infants start discovering the distributional properties and tracking the relationships between functors.

2.2 Cues related to prosody

2.2.1 Prosodic cues to speech segmentation: the role of rhythm

In the last decades, a great deal of research has focused on the role of prosody in the location of word boundaries in speech, revealing that both adults and infants make use of their native language's rhythmic properties to segment words from continuous speech (Cutler, Mehler, Norris and Seguí, 1986; Jusczyk, Houston and Newsome, 1999; Mehler, Dommergues, Frauenfelder and Seguí, 1981; Nazzi, Iakimova, Bertoncini, Frédonie, Alcantara, 2006). Rhythmic patterns vary cross-linguistically. Linguists have traditionally classified natural languages into three rhythmic classes, i.e., syllable-timed languages, stress-timed languages and mora-timed languages (Abercrombie, 1967; Ladefoged, 1975; Pike, 1945). The distinction between these three rhythmic classes results from their differing phonological properties (Dasher and Bolinger, 1982; Dauer, 1983; Ramus, Nespor and Mehler, 1999). Ramus, Nespor and Mehler (1999) conducted analyses of the acoustic properties of eight languages[3] and found that properties such as syllable structure (through measurement of consonantal intervals ΔC) and the proportion of vocalic intervals ($\%V$) led to a classification of the languages into the three rhythmic classes. Thus, stress-timed languages (e.g., English, Dutch), typically have a great variety of syllable types and tend to have heavy syllables, which results in increased duration of the syllable, entailing a lower vocalic ratio. Syllable-timed languages (e.g., Spanish, French) and mora-timed languages (e.g., Japanese), on the other hand, tend to have a more limited number of syllable types, hence having smaller consonantal intervals and a greater vocalic ratio than stress-timed languages.

As mentioned above, adult speakers exploit the rhythmic pattern of the native language to segment continuous speech into words. Cutler and Norris (1988), Mehler, Dommergues, Frauenfelder and Seguí, (1981) and Otake, Hatano, Cutler and Mehler, (1993) showed that speakers of languages that belong to different rhythmic classes make use of different segmentation strategies. Thus, speakers of syllable-timed languages (e.g., French, Spanish) segment speech at the boundaries of syllables, speakers of stress-timed languages (e.g., English, Dutch) segment speech at the onset of

[3] English, Polish, Dutch, French, Spanish, Italian, Catalan, and Japanese.

stressed syllables, and speakers of mora-timed languages (e.g., Japanese) segment speech at mora boundaries.

The pioneering work of Mehler, Dommergues, Frauenfelder and Seguí (1981) revealed that speakers of French—syllable-timed language—used the syllable as a unit of speech segmentation. In a target monitoring task, participants recognized faster target sequences (e.g., CVC*: pal*-, CV*: pa*-) embedded in bisyllabic nouns when the targets corresponded exactly to the first syllable of the nouns (e.g., CVC target: *pal*- in *pal-mier*, CV target: *pa*- in *pa-lace*) than when the targets did not correspond to a syllable (e.g., CV target: *pa*- in *pal-mier*, CVC target: *pal*- in *pa-lace*).

Cutler, Mehler, Norris and Seguí (1983, 1986) showed that the syllable was not the unit of speech segmentation for speakers of English, a stress-timed language. In a replica of Mehler et al.'s (1981) experiment with the original French stimuli, in addition to a similar experiment with English stimuli, English speakers did not show the syllabification effect obtained by Mehler et al. (1981) with speakers of French, i.e., English speakers had similar RTs to CV[C][4] (e.g., *ba*[*l*]-) and CVC targets (e.g., *bal*-) embedded in CV[C]- (e.g., *balance*) or CVC- words (e.g., *balcony*). Additionally, Otake, Hatano, Cutler and Mehler (1993) conducted an experiment similar to Mehler et al.'s (1981) with Japanese participants (i.e., speakers of a mora-timed language) and stimuli (e.g., targets: CV, *mo*- and CVC, *mon*-, stimuli: CV-, *monaka* and CVC-, *monka*), and found that, as the English speakers tested in Cutler et al., (1983, 1986), Japanese participants did not rely on the syllable to segment the input.

The syllable-based segmentation found in French speakers by Mehler et al. (1981) has been replicated in further studies on French (Pallier, Sebastián-Gallés, Felguera, Christophe and Mehler, 1993) and found with speakers of other syllable-timed languages, such as Spanish (Bradley, Sánchez-Casas and García-Albea, 1993; Sebastián-Gallés, Dupoux, Seguí and Mehler, 1992), Catalan (Sebastián-Gallés, Dupoux, Seguí and Mehler, 1992), Italian (Tabossi, Collina, Mazzetti and Zoppello, 2000), Portuguese (Morais, Content, Cary, Mehler and Seguí, 1989) and Korean (Kim, Davis and Cutler, 2008; Yoon and Derwing, 1995). Furthermore, speakers of syllable-timed languages implement this syllable-based segmentation strategy not only in the segmentation of input in the native language, but also when presented with input in a

[4] The brackets [ ] signal the ambisyllabicity of the consonant. English is characterized by the absence of clear syllabic boundaries and widespread ambisyllabicity.

foreign language, regardless of whether the foreign language belongs to the same rhythmic class as the native language, or to a different rhythmic class. Thus, French speakers show evidence of syllabification of stimuli in English (i.e., stress-timed language, Cutler et al., 1983, 1986) and Japanese (i.e., mora-timed language, Otake et al., 1993), and Korean speakers implement syllable-based segmentation when presented with French stimuli (Kim, Davis and Cutler, 2008).

Cutler and Norris (1988) showed that speakers of English segmented speech at the onset of strong syllables. Stress-timed languages are characterized by the opposition of strong (S) and weak (W) syllables. Strong syllables typically bear primary or secondary stress and contain full vowels. Meanwhile, weak syllables are unstressed and often contain short, reduced vowels like schwa.

Cutler and Norris (1988) found that when English speakers were asked to detect target words (e.g., *mint*) embedded in pairs of nonsense bisyllables, in which the target word spanned a syllabic boundary (e.g., *min-tesh*, *min-tayve*), they responded faster when the bisyllable had a SW structure than when the bisyllable had a SS structure. Cutler and Norris (1988) concluded that the English speakers had segmented the stimuli at the onset of strong syllables, which in the SS nonsense words led to delayed recognition of the target word split in the two syllables. Neither French nor Japanese speakers behave as English native speakers when presented with English stimuli (Cutler, Mehler, Norris and Seguí, 1983, 1986; Cutler and Otake, 1994).

Cutler and Norris (1988) proposed the Metrical Segmentation Strategy (MSS), according to which speakers of stress-timed languages treat strong syllables as initial syllables of lexical words, triggering segmentation. The efficiency of this stress-based strategy in English is backed up by Cutler and Carter's (1987) finding that 73% of all entries from an English dictionary analyzed had a strong initial syllable and 90% of the lexical words of a corpus analyzed started with a strong syllable. Further supporting evidence has been gathered in studies on word boundary misplacements and experiments on laboratory-induced misperceptions (Butterfield and Cutler, 1988; Cutler and Butterfield, 1992).

van Zon and de Gelder (1993) and Vroomen, van Zon and de Gelder (1996) showed that speakers of Dutch—a stress-timed language—also treated strong syllables as the initial syllables of lexical words, whereas weak syllables were treated as non-initial. Vroomen and de Gelder (1995) showed that, like in English, only 12.3% of all lexical words from a Dutch lexicon started with a weak syllable.

Otake, Hatano, Cutler and Mehler (1993) reported segmentation based on moraic units by speakers of Japanese —a mora-timed language. Participants detected CV targets (e.g., *ta*) embedded in CV*n*CV (e.g., *tanshi*) and CV*n*VCV (e.g., *tanishi*) stimuli equally easy, whereas CV*n* targets (e.g, *tan*) were harder to detect in CV*n*VCV words (e.g., *tanishi*) and elicited slower RTs in CV*n*CV (e.g., *tanshi*). These results are predicted by a mora-based segmentation strategy: CV targets corresponded to a single mora and were hence as easily detected in CV*n*VCV as in CV*n*CV words. CV*n* targets, however, were harder to detect in CV*n*VCV words, because in such words the second mora of the target (i.e., -*n*-) was the onset of the second syllable. Also, detection of CV*n* targets in CV*n*CV elicited long RTs because the listener had to match two morae instead of only one. Cutler and Otake (1994) and Otake, Hatano, Cutler and Mehler (1993) found that French and English speakers presented with Japanese stimuli did not implement the mora-based segmentation strategy as observed with speakers of Japanese.

Research on speech errors and word blending (Kubozono, 1989; 1996) have confirmed the role of the mora as the rhythmic unit in Japanese. Also, Cutler and Otake (1994) and Otake, Hatano and Yoneyama (1996) reported that speakers of Japanese tried to implement this mora-based segmentation when presented with English, Spanish and French stimuli.

Adult speakers rely thus on the rhythmic properties that characterize their native language to segment speech. The specific rhythm-based segmentation strategy implemented by speakers is determined by the rhythmic class to which their native language belongs.

A number of studies have examined whether bilingual speakers of languages that belong to different rhythmic classes develop the two rhythm-based segmentation procedures characteristic of the two languages they command, or if, on the contrary, bilinguals only implement one segmentation strategy (Cutler, Mehler, Norris and Seguí, 1989, 1992; Sanders, Neville and Woldorff, 2002).

Cutler, Mehler, Norris and Seguí (1989, 1992) examined the segmentation strategies of English/French early highly proficient bilinguals[5] both with English and French stimuli and found that bilinguals appeared to only be able to implement the segmentation strategy characteristic of their dominant language. The language of dominance was established by asking participants to choose which of the two languages they would prefer to keep if forced to choose one. Cutler et al. (1989, 1992) examined whether the bilinguals performed a syllable-based segmentation of Mehler et al.'s (1981) original materials in French and Cutler et al.'s (1986) original English materials. Also, Cutler et al. (1989, 1992) examined whether the French/English bilinguals performed a stress-based segmentation of the original English stimuli from Cutler and Norris (1988). Language dominance was the key factor in the analyses of their responses. The English-dominant bilinguals did not syllabify the French stimuli and used a stress-based segmentation of the English stimuli, i.e., they behaved as English monolingual speakers. French-dominant bilinguals, on the other hand, syllabified the French stimuli but did not use a stress-based segmentation of the English stimuli, behaving thus as French monolinguals. Also, the French-dominant bilinguals did not syllabify the English stimuli, unlike found in their monolingual counterparts (Cutler et al., 1983, 1986). Cutler et al. (1989, 1992) argue that this result suggests that high proficiency in the non-dominant language might lead to the inhibition of the inefficient strategy (i.e., syllabification of English stimuli).

van Zon (1997), as cited in Cutler (2001) provided supporting evidence for the bilinguals' inability to implement the rhythm-based segmentation strategy characteristic of the non-dominant language. This author presented French-dominant French/Dutch bilinguals with Dutch stimuli and found that the bilinguals did not use stress-based segmentation of the stimuli. However Sanders, Neville and Woldorff (2002) found stress-based segmentation by L1Japanese/L2English and L1Spanish/L2English late, highly proficient bilinguals with English stimuli. In a phoneme detection task, both groups were significantly more accurate in spotting the target phoneme in words (embedded in sentences) that had the stress pattern characteristic in English (i.e., strong-initial), than in words with the infrequent stress pattern (i.e., weak-initial). Sanders et al. (2002) concluded that these late bilinguals seemed to have learnt the segmentation cue characteristic of their L2. Note however that though Sanders et al. (2002) reported the

[5] Most of the participants had been raised with in a bilingual environment, i.e., with a French-speaking parent and an English-speaking parent.

predominant use of English by both groups the language of preference was not reported, and therefore dominance was not accounted for in similar terms to Cutler et al.'s (1989, 1992) investigation. These conflicting results on the bilinguals' ability to implement the rhythm-based segmentation strategies of the bilinguals' two languages reveal the need for further research.

Nazzi, Iakimova, Bertoncini, Frédonie and Alcantara (2006) proposed that the early segmentation abilities in infants follow a rhythm-based strategy similar to the one reported for adults. According to this rhythmic segmentation hypothesis, infants' early segmentation procedures differ depending on the global rhythmic properties of their native languages. Indeed, rhythm is a property of language acquired very early by infants. Speech rhythm is audible in the womb, and Condon and Sander (1974) reported synchronies between infants' movements and speech input. 2- to 5-month-old infants, newborns and even near-term foetuses are able to discriminate between their native language and a foreign language that belongs to a different rhythmic class, even when low-pass filtered to 400 Hz, hence eliminating segmental information (Bosch and Sebastián-Gallés, 1997; Christophe and Morton, 1998; Dehaene-Lambertz and Houston, 1998; Kisilevsky et al., 2009; Mehler, et al., 1988; Nazzi, Jusczyk and Johson, 2000). Furthermore, newborns discriminate between two foreign languages that belong to different rhythmic classes, also when low-pass filtered to 400 Hz (Mehler et al., 1988[6]; Nazzi et al., 1998; Ramus, 2002; Ramus et al., 2000). However, newborns do not discriminate between two foreign languages that belong to the same rhythmic class (Nazzi et al., 1998).

Infants' ability to discriminate the native language from a foreign language that belongs to the same rhythmic class seems to undergo a change during the first months of age. Thus, 2-month-olds do not behave homogenously and show only marginally significant discrimination (Christophe and Morton, 1998), while by 4 months of age infants accomplish discrimination between the native language and a language that belongs to the same class (Bosch and Sebastián-Gallés, 1997, 2001), and by 5 months of age infants discriminate two varieties of the native language (Nazzi, Jusczyk and Johnson, 2000). Regarding the discrimination abilities of bilingual infants, Bosch and

[6] Mehler et al. (1988) initially report no discrimination by 4-day-old French newborns between English (stress-timed) and Italian (syllable-timed), but a reanalysis is carried out in Mehler and Christophe (1995) that reveals a significant difference.

Sebastián-Gallés (1997) reported that 4-month-old infants raised in a bilingual Catalan/Spanish environment discriminated between their native languages and a foreign language that belonged either to the same rhythmic class (i.e., Italian), or to a different class (i.e., English), and Bosch and Sebastián-Gallés (2001) showed that 4-month-olds infants discriminated between the two native languages (i.e., Catalan and Spanish).

In parallel to this development, infants around 6 months of age start to display a preference for the rhythmic pattern typical in their native language. German-, Dutch- and English-learning 6- to 10-month-olds prefer the trochaic pattern, i.e., the predominant rhythmic pattern in stress-timed languages characterized by alternation of strong and weak syllables, and treat strong syllables as the onset of new words (Höhle, 2002; Höhle, Bijleljac-Babic, Herold, Weissenborn, Nazzi, 2009; Jusczyk, Cutler and Redanz, 1993; Kooijman, 2007; Morgan, 1996; Turk, Jusczyk and Gerken, 1995). Infants as young as 7½ months of age segment bisyllabic words with a trochaic pattern (i.e., SW) from fluent speech, but fail to segment words with an iambic pattern (i.e., WS) (Jusczyk, Houston and Newsome, 1999). Additionally, French-learning 12-month-old infants segment speech into syllable-sized units (Goyet, de Schonen and Nazzi, 2010; Nazzi et al., 2006). Although research on mora-based segmentation by infants is still pending, studies focused on Japanese infants' sensitivity to mora-quantity have reported discrimination of stimuli based on mora-quantity in 8- to 12-month-olds (Mazuka and Hayashi, 2006, as cited in Mazuka, 2007; Mugitani, Kobayashi and Amano, 2005, as cited in Mazuka, 2007; Sato, Sogabe and Mazuka, 2010).

Infants seem to be born equipped with a general purpose mechanism that allows them to detect variations between the three rhythmic classes found in the studies with adults: syllable-timed, stress-timed and mora-timed.

2.2.2 Prosodic cues to phrase segmentation

In parallel to the investigation on prosodic cues to word segmentation discussed in the previous section, a great deal of research has focused on the role of prosody in the segmentation of larger syntactic constituents from speech, such as phrases and clauses. It has long been noted that no one-to-one correlation exists between syntactic and prosodic constituents. Chomsky and Halle (1968: 372) illustrate this non-isomorphism between prosodic and syntactic bracketing of sentences with the following example:

Example 1. Non-isomorphism between prosodic and syntactic bracketing

(1) Syntactic bracketing
This is [the cat that caught [the rat that stole [the cheese]]]

(2) Prosodic bracketing
[This is the cat] [that caught the rat] [that stole the cheese]

Prosodic constituents are organized into a phonological hierarchy, as shown by Nespor and Vogel (1982, 1986) and Selkirk's (1978, 1980, 1986, 1996) research on the domains of intonational contours and phonological rules. However, prosodic constituents above the prosodic word level usually align with the limits of major syntactic constituents. Selkirk (1996, 2000) captures the alignment of phonological and syntactic phrases with the *Align-XP* constraint, which states that the right or left edge of a syntactic phrase (XP) must be aligned with the right or left edge of a phonological phrase. Intonational and phonological phrases (two of the members of the phonological hierarchy) tend to align with clause and phrase boundaries (Cruttenden, 1986; Nespor and Vogel, 1986; Selkirk, 1996).

Research on the acoustic properties of clause and phrase boundaries have shown that these boundaries are frequently marked by pauses, final lengthening and changes in fundamental frequency (Cooper and Paccia-Cooper, 1980, as cited in Jusczyk et al., 1992; Jusczyk, Hirsh-Pasek, Kemler Nelson, Kennedy, Woodward and Piwoz, 1992; Scott, 1982). Additionally, these cues to clause and phrase boundaries appear to be exaggerated in infant-directed speech (Fisher and Tokura, 1996).

Christophe and Dupoux (1996) claim that "*speech is spontaneously perceived as a string of prosodic constituents* [...]" (p.395) and that adults and infants make use of the available phonological cues in the input to segment speech. Gleitman and Wanner (1982) propose that this information might allow infants to locate boundaries to syntactic constituents prior to lexical knowledge. This proposal is known as the phonological bootstrapping hypothesis, or prosodic bootstrapping hypothesis. Furthermore, also within the framework of the phonological bootstrapping hypothesis, Morgan and Demuth (1996) and Nespor, Guasti and Christophe (1996) claim that the prosodically-driven division of speech into smaller chunks might allow infants to detect syntactic regularities, which in turn enable them to build rudimentary representations of certain syntactic distinctions and rules. Christophe and Dupoux (1996) suggest that the segmentation of speech into smaller units would reduce the workload required by other distributional or phonological processes (e.g., phonotactic constraints). Thus, phonological bootstrapping is not proposed as the sole source of information leading to the acquisition of syntactic structures, but as a mechanism that works in parallel to other bootstrapping mechanisms, e.g., distributional learning (Morgan and Demuth, 1996).

It has been proposed that the prosodic cues present in the input may allow infants to bootstrap the basic word order of the language under acquisition, setting the values of major syntactic parameters such as the Head-Directionality parameter or the Branching-Direction parameter prior to lexical knowledge (Christophe, Guasti, Nespor, Dupoux and van Ooyen, 1997; Mazuka, 1996; Nespor, Guasti and Christophe, 1996; Nespor, Shukla, van de Vijver, Avesani, Schraudolf and Donati, 2008).

Infants are sensitive to the word order of the language under acquisition before their first multi-word utterances, as shown by Hirsh-Pasek, Golinkoff, Fletcher, DeGaspe-Beaubien and Cauley (1985), as cited in Mazuka (1996). These authors presented infants in the one-word stage with utterances such as *Oh look! Cookie Monster is tickling Big Bird*. Infants saw two monitors: one monitor played a video that matched the utterance (i.e., Cookie Monster tickling Big Bird), while another monitor played a video that presented the reverse action (i.e., Big Bird tickling Cookie Monster). The infants preferred to look at the video that matched the utterance, than to the video that showed the reverse action.

Mazuka (1996) suggested that infants might set the Branching-Directionality parameter during the first year of life, by means of prosodic information. This parameter refers to the direction of recursive embedded clauses relative to the main clause. In right-branching languages (e.g., English) embedded clauses appear to the right of the main clause, whereas in left-branching languages (e.g., Japanese) embedded clauses appear to the left of the main clause (Mazuka, 1996). Mazuka (1996) proposes that the input might contain reliable cues that signal the dominance relation between main and embedded clauses, which might in turn, allow infants to discover the branching direction of the language under acquisition.

Additionally, the location of prominence within phonological phrases has been proposed as a potentially useful cue that might lead infants to set the value of the Head-Directionality parameter, due to its correlation with the relative order of heads and complements in natural languages (Christophe, Guasti, Nespor, Dupoux and van Ooyen, 1997; Nespor, Guasti and Christophe, 1996; Nespor, Shukla, van de Vijver, Avesani, Schraudolf and Donati, 2008). This proposal will be discussed in detail in Chapter 5.

Infants' acquisition of prosodic cues to syntactic constituency appears to develop gradually, i.e., prosodic cues to clauses are detected earlier in development than prosodic cues to phrases, which suggests that infants might require greater familiarity with the prosodic structure of the language in order to detect the prosodic cues to phrases (Jusczyk, Hirsh-Pasek, Kemler Nelson, Kennedy, Woodward and Piwoz, 1992). Hirsh-Pasek, Kemler Nelson, Jusczyk, Wright Cassidy, Druss and Kennedy (1987) showed that English-learning 7- to 10-month-old infants listened longer to speech samples of infant-directed speech in which 1 second pauses were inserted at clause boundaries, than to samples in which pauses were inserted between words within a clause. Jusczyk, Hohne and Mandel (1995) replicated this preference with 4½- and 6-month-old infants and with low-pass filtered versions of the materials with a new group of 6-month-olds. Furthermore, English-learning 4½- but not 6-month-old infants discriminated similar stimuli in Polish, that is, in a foreign language. Jusczyk et al. (1995) argued that this finding suggested that infants might initially pick up general prosodic cues that are similar across-languages and gradually tune to the specific characteristics of the native language.

By 6 months of age, infants make use of prosodic cues to syntactic clauses to divide the input. Nazzi, Kemler Nelson, Jusczyk and Jusczyk (2000) familiarized 6-month-old infants with repetitions of short sequences of words that consisted of either a prosodically well-formed clause (e.g., *Leafy vegetables taste so good*.) or a prosodically ill-formed clause that began and ended in the middle of two adjacent clauses (e.g., *leafy vegetables. Taste so good*). The infants listened longer to passages that contained the familiarized prosodically well-formed sequences than to passages containing the familiarized but prosodically ill-formed sequences.

The use of prosodic cues to syntactic phrases in speech segmentation seems to emerge around the second half of the first year of life, the time at which perceptual reorganization takes place (Bosch and Sebastián-Gallés, 2003; Polka and Werker, 1994; Werker and Lalonde, 1988; Werker and Tees, 1984). Thus, 6- to 9-month-old infants make use of prosodic cues to syntactic phrases in speech segmentation, albeit this ability appears to be in a developing stage and crucially depends on the quantity and strength of the prosodic cues that signal the phrases (Jusczyk, Hirsh-Pasek, Kemler Nelson, Kennedy, Woodward and Piwoz, 1992; Soderstrom, Seidl, Kemler Nelson and Jusczyk, 2003). Soderstrom, Seidl, Kemler Nelson and Jusczyk (2003) showed that 6- and 9-month-old infants listened longer to passages that contained a previously familiarized well-formed phrase than to a familiarized and phonemically identical sequence that straddled a phrase boundary. However, this preference was not replicated with a second set of stimuli. Soderstrom et al. (2003) conducted acoustic analyses on the two sets of stimuli, which revealed that the first set of stimuli contained more acoustic cues that additionally were more marked than the second set of stimuli. Infants' sensitivity to prosodic cues to syntactic phrases, i.e., phonological phrases, will be discussed in greater detail in Chapter 5.

Adult speakers, like infants, are sensitive to the prosodic markers to syntactic constituents. Christophe, Peperkamp, Pallier, Block and Mehler (2004) showed that phonological phrase boundaries constrained lexical access in French adult speakers. In a word-monitoring task, these authors found that activation of lexical competitors (e.g., *chagrin*) to a target word (e.g., *chat*) was constrained when the competitor straddled a phonological phrase boundary (e.g., [*D'après ma sœr*], [*le gros chat*] [*grimpait aux arbres*]...), but not when the competitor straddled a prosodic word boundary within a single phonological phrase (e.g., [*Le livre*] [*racontait l'histoire*] [*d'un chat grincheux*]...).

Also, artificial language learning studies with adults have shown that prosodic cues facilitate segmentation of an artificial language and learning phrase structure (Bagou, Fougeron and Frauenfelder, 2002; Langus, Marchetto, Bion and Nespor, 2012; Morgan, Meier and Newpor, 1987). Thus, Bagou, Fougeron and Frauenfelder (2002) showed that prosodic information (i.e., changes in duration and pitch) helped participants to segment sequences with variable lengths from an artificial language. Langus, Marchetto, Bion and Nespor (2012) presented participants with an artificial language that contained two types of prosodic cues signalling two different levels of syntactic constituency. These cues were final lengthening at the boundaries of phrase-like units and pitch declination in sentence-like units. Additionally, the artificial language contained non-adjacent dependencies at the phrase-like and sentence-like levels. Participants were able to learn both the “phrase” and “sentence” level rules of the artificial language, unlike participants in a subsequent experiment with a structurally similar but prosodically flat artificial language.

Adults and infants are thus sensitive to the prosodic cues to major syntactic constituents present in the signal. Infants make use of these cues to segment the input prior to lexical knowledge. Moreover, these prosodic cues might help infants bootstrap the acquisition of certain syntactic properties of the language. Similarly, prosodic information constrains lexical access in adults and enables learning rules in artificial languages.

2.3 Statistical and prosodic cues in a hierarchical account of segmentation cues

The previous sections focused on the role of statistical and prosodic cues in speech segmentation. However, as mentioned in Chapter 1, no single cue suffices to achieve accurate segmentation of speech (Christiansen, Allen and Seidenberg, 1998). For instance, rhythm-based segmentation (as implemented by the Metrical Segmentation Strategy in English) leads to missegmentation of words with an initial weak syllable (e.g., misperception of the word *analogy* as *and allergy*, Cutler and Butterfield, 1992). Similarly, computation of TPs may also lead to missegmentation (e.g., a children’s reply “*I am have!*” when asked to behave, as a result of the overall high frequency of the syllable *be*, Peters, 1985, as cited in Johnson and Jusczyk, 2001). Successful word segmentation depends thus on the listener’s ability to integrate the available cues (Christiansen, Allen and Seidenberg, 1998).

In addition to prosodic and statistical cues, adults and infants use other prelexical cues in speech segmentation, such as phonotactic constraints and acoustic-phonetic cues (e.g., allophonic variation and coarticulation). Also, adult speakers make use of lexical-semantic information in word recognition when listening to the native language, but this information is unavailable or scarce for infants during acquisition (Mattys, White and Melhorn, 2005).

Phonotactic constraints refer to the language-specific restrictions on the permissible phonemes and sequences of phonemes both within- and between-words (McQueen, 1998). As mentioned in section 2.1.1, phonotactics specify (a) categorical constraints, which refer to the legal and illegal phoneme sequences in a given language, or (b) probabilistic constraints, which refer to the relative frequency of occurrence of the legal phoneme sequences in a given language. It has been proposed that knowledge of the native language's phonotactic constraints provides the listener with cues to word boundaries, i.e., a phonotactically illegal sequence entails the presence of a boundary (Brent and Cartwright, 1996; McQueen, 1998). Similarly, a low-frequency sequence is more likely to contain a word boundary than a high-frequency sequence. Computational simulations have revealed that phonotactic constraints are a potentially powerful cue to word segmentation (Brent and Cartwright, 1996; Cairns, Shillcock, Chater and Levy, 1997).

Adult speakers' sensitivity to both categorical and probabilistic phonotactic constraints of the native language has been well established in numerous studies (Dehaene-Lambertz et al., 2000; Dupoux, Kakehi, Hirose, Pallier and Mehler, 1999; Fais, Kajikawa, Werker and Amano, 2005; Massaro and Cohen, 1983; van der Lugt, 2001; Vitevitch, Luce, Charles-Luce and Kemmerer, 1997; Weber, 2001). Massaro and Cohen (1983) showed that categorization of an ambiguous consonantal stimulus embedded in a CCV syllable was influenced by the phonotactic legality of the resulting cluster. Moreover, word-spotting and phoneme monitoring studies have shown that adult speakers make use of both categorical and probabilistic phonotactics in speech segmentation (Dumay, Frauenfelder and Content, 2002; McQueen, 1998; van der Lugt, 2001; Warner, Kim, Davis and Cutler, 2005; Weber, 2001; Yip, 2000, as cited in Yip, 2006). Thus, Weber (2001) conducted a phoneme monitoring task with speakers of German and found that participants detected the target phoneme (e.g., /*t*/) slower when the phoneme appeared in a phonotactically legal context (e.g., [*ʃtɪm*]) than when the phoneme occurred in an illegal context (*[*stɪm*]).

Sensitivity to phonotactic constraints in infants appears around 9 months of age (Friederici and Wessels, 1993; Friedrich and Friderici, 2005; Jusczyk, Friederici, Wessels, Svenkerud and Jusczyk, 1993; Kajikawa, Fais, Mugitani, Werker and Amano, 2006; Mugitani, Fais, Kajikawa, Werker and Amano; Sebastián-Gallés and Bosch, 2002). Friederici and Wessels (1993) showed that Dutch-learning 9-month-old infants preferred to listen to stimuli that followed the phonotactic constraints of their native language (e.g., *bref*), rather than to stimuli that contained illegal phonotactic sequences (e.g., *rtum*). Furthermore, by 9 months of age, infants have developed detailed knowledge of the frequency of appearance of phonotactic patterns in their native language and prefer to listen to words that contain phoneme sequences that are highly frequent in their native language, rather than to listen to less frequent sequences (Jusczyk, Luce and Charles-Luce, 1994; Mattys, Jusczyk, Luce and Morgan, 1999). Also, 9-month-old infants use both categorical and probabilistic phonotactic knowledge to locate word boundaries in fluent speech (Friederici and Wessels, 1993; Mattys and Jusczyk, 2001).

On a par with phonotactic constraints, allophonic variation assists infants and adults in speech segmentation. Allophones are the context-dependent variants of a given phoneme. The distribution of allophones is constrained to a certain position within a word or even across words, in the continuous speech signals. As noted by Kahn (1976) and Lisker and Abramson (1964), the aspirated variant [t^h] of the phoneme /*t*/ in English occurs as the onset of stressed syllables, and therefore is usually found in word-initial position (e.g., *tap, Tom*), while the unaspirated allophone [*t*] is found in other positions (e.g., *butter, cat, atlas*). Correlations between allophonic distribution and word boundaries have been found in studies on English (Lehiste, 1960; Nakatani and Dukes, 1977). Hence, allophonic variation has been proposed as a useful cue to the detection of word boundaries (Church, 1987), e.g., the presence of the aspirated allophone [t^h] of the phoneme /*t*/ in a sequence in English would probably indicate the presence of a preceding boundary.

Adult speakers are sensitive to the allophonic information present in the input and make use of this cue to segment speech (Lehiste, 1960; Nakatani and Dukes, 1977;

Smith and Hawkins, 2000; Whalen, Best and Irwin, 1997). Smith and Hawkins (2000) embedded English target words (e.g., *reap, bell*) in pairs of nonsense words. Both members of each pair were identical except for allophonic variation of a segment of the target word (e.g., *reap* in [*gəɹɪp*] vs. [*gədɹ̥ɪp*]. The allophone in one member of each pair (e.g., [*r*] in [*gəɹɪp*]) was appropriate for the target word in isolation (e.g., *reap:* [*ɹɪp*]), while the other allophone was inappropriate (e.g., [*ɹ̥*] in [*gədɹ̥ɪp*] for *reap*: *[*ɹ̥ɪp*]). Participants spotted the target word slower and were less accurate when the stimuli contained inappropriate allophones than when the stimuli contained appropriate allophones.

Sensitivity to allophonic contrasts has been attested in infants as young as 2 months of age (Hohne and Jusczyk, 1994). Hohne and Jusczyk (1994) reported that English-learning 2-month-old infants discriminated between two phonemically identical items (i.e., *nitrate* vs. *night rate*) that differed in the allophonic variation of the two phonemes /*t*/ and /*r*/. However, the use of allophonic cues in word segmentation from fluent speech appears around 10½ months of age (Jusczyk, Hohne and Bauman, 1999). Thus, Jusczyk, Hohne and Bauman (1999) showed that English-learning 10½-month-olds familiarized with repetitions of one member of the *nitrate* vs. *night rate* pair created by Hohne and Juscyzk (1994) listened longer to passages that contained the familiarized stimulus than to passages containing the non-familiarized member of the *nitrate* vs. *night rate* pair.

The input signal contains coarticulatory information, which is picked up by adults and infants and used to locate word boundaries in speech. Coarticulation refers to the overlap of adjacent articulations, that is, when the acoustic properties of a phonetic segment blend with the acoustic properties of a surrounding segment (Johnson and Jusczyk, 2001; Ladefoged, 1975). Coarticulation can be both anticipatory, i.e., where a phonetic segment assumes a feature from a following segment (e.g., vowel nasalization followed by a nasal consonant), or perseverative, i.e., where a phonetic segment assumes a feature from a preceding segment (e.g., /*g*/ is fronted when preceded by /*l*/ in English). The degree of coarticulation is greater between segments that occur within words than between segments that span a word boundary (Ladefoged, 1975).

Adult listeners are sensitive to the coarticulatory cues present in the input and make use of this information to segment words from speech (Fernandes, Ventura and Kolinsky, 2007; Mann, 1980; Mattys, 2004; Mattys, White and Melhorn, 2005). Thus, Mattys, White and Melhorn (2005) conducted a lexical decision task using the cross-modal priming fragment paradigm with speakers of English. Participants listened to nonsense utterances (e.g, /*revəmətɪ*/) which contained bisyllabic primes (e.g., /*mətɪ*/ as in *material*). The primes were either coarticulated with the preceding context, disfavouring boundary insertion prior to the prime, or were decoarticulated with the context, favouring the presence of a boundary at the onset of the target. After each utterance, participants saw a string of letters on a monitor (e.g., *material*) and had to decide whether the letter string constituted a word. Priming was facilitated by favourable coarticulatory cues, i.e., adults made use of coarticulatory information to segment the primes from the nonsense utterances.

Infants' sensitivity to at least some kinds of coarticulatory information appears to emerge at around 4 to 5 months of age (Fowler, Best, McRoberts, 1990). Around 7 to 8 months of age, infants are able to locate word boundaries relying on coarticulatory information (Curtin, Mintz and Byrd, 2001; Johnson and Jusczyk, 2001). Johnson and Jusczyk (2001) familiarized English-learning 8-month-olds to a 2½ minute long artificial language that consisted of concatenations of four trisyllabic items (e.g., *golatu*, *daropi*). The last syllable of some trisyllabic items was coarticulated to the first two syllables of the following item (e.g., *gola<u>tudaro</u>pi*), creating trisyllabic part-items (e.g., *tudaro*). After exposure, the infants listened to repetitions of the original trisyllabic items (e.g., *golatu*) and decoarticulated versions of the coarticulated part-items heard in the familiarization (e.g., *tudaro*). Infants listened longer to the items than to the decoarticulated part-items, but a control experiment showed that this preference revealed a novelty effect, i.e., infants had segmented the coarticulated sequences from the familiarization.

In addition to the above discussed prelexical cues, lexical information is also an available cue to adult speakers. It has been proposed that two processes guide adult word recognition: multiple lexical activation and competition between word candidates (Luce and Pisoni, 1998; Marslen-Wilson, 1987; McClelland and Elman, 1986; Norris, 1994). Thus, studies with lexical decision tasks that used the cross-modal priming paradigm have shown that, at the early stages of word recognition, multiple lexical candidates that match the input are simultaneously activated (Marslen-Wilson, 1987, Zwitserlood, 1989). These lexical candidates compete with each other and are inhibited until only one candidate is selected (McQueen, Norris and Cutler, 1994). Thus, McQueen, Norris and Cutler (1994) showed that participants spotted a target word (e.g., *mess*) slower when the target was embedded in a bisyllable that matched with the onset of an English word (e.g., /*dəmɛs*/ as in *domestic*) than when the target occurred in a bisyllable that did not match with the onset of an English word (e.g., /*nəmɛs*/). These results show that the lexical candidate *domestic* was simultaneously activated and in competition with the target.

As mentioned at the beginning of this section, word segmentation results from the integration of the various available cues present in the input, rather than by the information provided by one single cue. Adult speakers integrate lexical and prelexical cues during speech processing (Dumay, Content and Frauenfelder, 1999; McQueen, Norris and Cutler, 1994; Norris, McQueen and Cutler, 1995; Sanders and Neville, 2000; Smith and Hawkins, 2000). Infants start integrating the two earliest available cues (i.e., statistical information and prosodic information) at around 8 to 9 months of age. Speech segmentation is enhanced when these two cues act simultaneously (Morgan, 1994; Morgan and Saffran, 1995). Acquisition of other cues that takes place later in the development (e.g., phonotactics, allophonic variation) results in more fine-grained word segmentation abilities. For instance, English-learning 10½-month-olds but not 7½-month-olds segment words with the non-predominant rhythmic pattern (weak-strong, WS) in English from fluent speech (Jusczyk, Houston and Newsome, 1999). By 10½ months of age infants appear thus to have acquired other segmentation cues (e.g., phonotactic constraints) that enable them to segment WS words from speech.

Research has tried to disentangle the relative weight that adults and infants give to each segmentation cue. To that end, several studies have pitted segmentation cues against each other, and the results obtained have revealed that not all cues are equally weighed and that the relative weigh given to these cues changes throughout development (Fernandes, Ventura and Kolinsky, 2007; Johnson and Jusczyk, 2001; Mattys, Jusczyk, Luce and Morgan, 1999; Mattys, White and Melhorn, 2005).

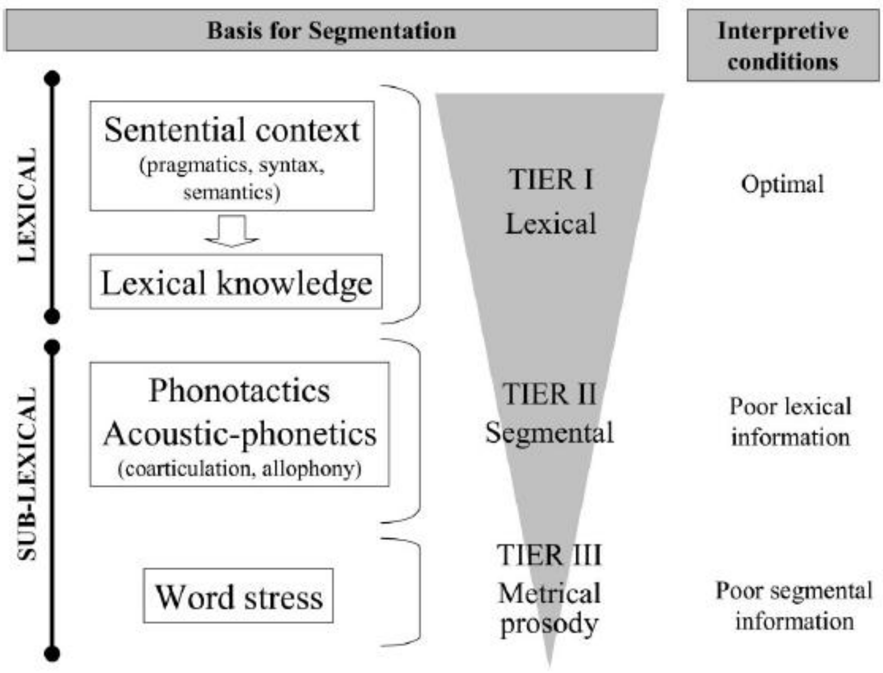
Speech segmentation in adult speakers appears to be primarily driven by lexical information, though modulated by phonotactic, acoustic-phonetic (i.e., allophonic, coarticulatory) and rhythmic cues (Andruski, Blumstein and Burton, 1994; Dumay, Content and Frauenfelder, 1999; McQueen, Norris and Cutler, 1994; Norris, McQueen and Cutler, 1995; Sanders and Neville, 2000; Smith and Hawkins, 2000; Vroomen, van Zon and de Gelder, 1996). When pitted against each other, lexical information outranks conflicting rhythmic, phonotactic and coarticulatory cues (Mattys, White and Melhorn, 2005). Additionally, segmental cues (e.g., phonotactic, coarticulation) outweigh rhythmic and statistical cues (Fernandes, Ventura and Kolinsky, 2007; Finn and Hudson Kam, 2008; Mattys, White and Melhorn, 2005; Mersad and Nazzi, 2011). Last, prosodic cues are more heavily weighed than statistical cues (Langus, Marchetto, Bion and Nespor, 2012; Shukla, Nespor and Mehler, 2007).

In a series of word-spotting and lexical decision tasks with English speakers, Mattys, White and Melhorn (2005) pitted lexical information against rhythmic, coarticulatory and phononotactic cues. They found that adult speakers relied on lexical cues rather than on rhythmic or segmental cues (i.e., convergent coarticulatory and phonotactic cues), but made use of segmental and rhythmic cues whenever lexical information was impoverished. In turn, Mattys et al. (2005) pitted rhythmic cues against segmental cues (i.e., coarticulation, phonotactics), and showed that listeners weighed segmental cues more heavily than rhythmic cues. Whenever the signal was degraded, rendering segmental cues unavailable, listeners relied on the rhythmic information present in the input. Fernandes, Ventura and Kolinsky (2007) pitted statistical cues (TPs) against coarticulatory cues in an artificial language learning experiment with speakers of Portuguese and found prevalence of segmental cues over statistical information. Finn and Hudson Kam (2008) reported similar prevalence of segmental cues over TPs in a series of three artificial language learning experiments with English speakers, i.e., the native language's phonotactic constraints hindered the extraction of statistically defined word-like items from the artificial languages. Thus, participants failed to segment the word-like items from the artificial language whenever these items yielded phonotactically illegal sequences in English.

Langus, Marchetto, Bion and Nespor (2012) and Shukla, Nespor and Mehler (2007) showed that adult speakers weigh prosodic cues more heavily than statistical cues. Thus, Shukla, Nespor and Mehler (2007) found that Italian adult speakers that were familiarized in an artificial language that contained prosodic contours (intonational phrases or IPs) extracted from either Italian or Japanese natural utterances, did not recognize statistically defined (TPs) word-like items that straddled IP boundaries, but did recognize statistically defined word-like items that occurred within an IP during familiarization. Similarly, Langus, Marchetto, Bion and Nespor (2012) showed that adult speakers relied on prosodic cues in the segmentation of an artificial language that contained conflicting prosodic and statistical cues and that participants only segmented the input based on TPs between adjacent syllables in the absence of the conflicting prosodic cues.

Based on the evidence presented above, Mattys, White and Melhorn (2005) propose a hierarchical account of segmentation cues, which claims that these cues are organized into a three tiered structure, depending on the relative weight speakers give to each cue:

Figure 3. Hierarchy of segmentation cues. From Mattys, White and Melhorn (2005:488)



According to this model, adults rely primarily on lexical information (tier I) to segment words when the signal contains all available cues optimally. When lexical knowledge is unavailable (e.g., impoverished or ambiguous), adults make use of segmental cues (tier II: phonotactics, allophones, coarticulation). Adults rely on rhythmic cues (tier III) as a last resort when the signal contains poor segmental information (e.g., in a noise-degraded signal), as shown by Cutler and Butterfield (1992) and Smith, Cutler, Butterfield and Nimmo-Smith (1989). Though Mattys, White and Melhorn (2005) do not include statistical information in the hierarchy, the results by Langus, Marchetto, Bion and Nespor (2012) and Shukla, Nespor and Mehler (2007) showing that prosodic cues outrank statistical information suggest that statistical information forms part of a fourth tier in the hierarchy of segmentation cues.

Research on the acquisition and relative weight of segmentation cues in infants points to a bottom-up access to the hierarchy at the onset of language acquisition (Johnson and Jusczyk, 2001; Johnson and Seidl, 2009; Mattys, Jusczyk, Luce and Morgan, 1999; Mattys, White and Melhorn, 2005; Thiessen and Saffran, 2003). Infants' lexical knowledge during the earlier stages of acquisition is very scarce at best. Acquisition of segmental cues takes place at around the second half of the first year of life (Curtin, Mintz and Byrd, 2001; Friederici and Wessels, 1993; Jusczyk, Hohne and Bauman, 1999; Mattys and Jusczyk, 2001), whereas statistical and prosodic information are the earliest segmentation cues used by infants (Johnson and Tyler, 2010; Jusczyk, Houston and Newsome, 1999).

Several studies have pitted cues against each other in order to establish the development of the infants' segmentation skills. Initially, word segmentation appears to be primarily driven by statistical cues. Thus, Thiessen and Saffran (2003) pitted TPs against rhythmic cues in an artificial language learning experiment with English-learning 6½- to 7-month-olds and found that infants relied more heavily on statistical cues than on prosodic cues. However, by 8 months of age TPs are outranked by rhythmic cues (Johnson and Jusczyk, 2001). Prevalence of rhythmic over statistical cues has been also attested in 9- and 11-month-old infants (Johnson and Seidl, 2009; Thiessen and Saffran, 2003). These results tally with the results obtained with adult speakers, and suggest that statistical learning belongs to a lower tier than prosodic cues within the hierarchy.

Studies that pitted segmental cues against prosodic and statistical cues are scarce (Johnson and Jusczyk, 2001; Mattys, Jusczyk, Luce and Morgan, 1999). Johnson and Jusczyk (2001) reported that coarticulatory information outweighed statistical information (TPs) in English-learning 8-month-old infants. Thus, the 8-month-olds relied on coarticulatory cues to segment items from an artificial language that conflicted with statistical information. Also, Mattys, Jusczyk, Luce and Morgan (1999) presented

English-learning 9-month-olds with CVC.CVC bisyllables with either SW or WS pattern that contained mismatching phonotactic cues. Phonotactic constraints signalled the presence of a boundary within the CC cluster (i.e., CVC.CVC) in SW stimuli, whereas phonotactics disfavoured the presence of a boundary in WS stimuli. The 9-month-olds listened longer to the SW than to the WS stimuli, i.e., rhythmic cues overrode phonotactic cues.

From the two earliest available sources of information at the onset of language acquisition, infants appear to weigh statistical information more heavily than rhythmic cues, though statistical information is rapidly overridden by rhythmic and segmental cues (i.e., coarticulatory information). Thiessen and Saffran (2003) propose that statistical cues (TPs) may help infants discover the rhythmic pattern of the language under acquisition. However, computational modelling of infant-directed speech has obtained conflicting results: Swingley (2005) provided supporting evidence that statistical information can help infants extract word forms that correspond to the rhythmic pattern of the native language, whereas Gambell and Yang (2004) and Yang (2004) found that analysis based on TPs yielded very poor results, due to, for instance, the abundance of monosyllables in English.

Though by 9 months of age infants already have acquired the phonotactic constraints of the native language, rhythmic cues are still more heavily weighed. The fact that segmental information outranks statistical information in 8-month-olds but rhythmic information still outranks segmental information in 9-month-olds suggests that rhythmic cues might play a more important role than statistical information in speech segmentation. Research aimed to establish the age around which segmental cues override rhythmic cues is pending. The information provided by these prelexical cues allows infants to start building a lexicon that leads to access to the first tier of the hierarchy.

Chapter 3. Exploring segmentation preferences in monolinguals and bilinguals

3.1 Introduction

Discovering the word order of the target language is one of the challenges that infants face during acquisition. The configuration of the basic word order appears to take place very early in the development: infants' first multiword utterances follow the word order rules of the target language (Brown, 1973, as cited in Guasti, 2002). As mentioned in Chapter 2, it has been proposed that the input contains cues that might allow infants to segment the continuous material that they hear, discover regularities and bootstrap the acquisition of syntax (Gervain, 2007; Gleitman and Wanner, 1982, Morgan and Demuth, 1996). These bootstrapping hypotheses suggest that infants might set the value of the Head-Directionality parameter (Baker, 2001; Chomsky, 1981) prior to lexical knowledge, by means of prosodic cues (i.e., phonological bootstrapping hypothesis) or statistical cues (i.e., frequency-based bootstrapping hypothesis) (Christophe, Guasti, Nespor, Dupoux and van Ooyen, 1997; Gervain, 2007; Nespor, Guasti and Christophe, 1996; Nespor, Shukla, van de Vijver, Avesani, Schraudolf and Donati, 2008).

Gervain (2007) proposes a frequency-based bootstrapping hypothesis, according to which the high frequency of functors and their relative order of appearance with respect to content words in natural languages are useful cues that enable infants to segment the input into phrases and set the value of the Head-Directionality parameter, bootstrapping a rudimentary representation of word order, i.e., an abstract property of the target language. The order of functors and content words correlates with the order of heads and complements in natural languages (Greenberg, 1963). Thus, in head-initial (VO) languages functors typically occur phrase-initially, whereas in head-final (OV) languages functors usually appear phrase-finally.

Gervain (2007) showed that adult speakers track the relative distribution of frequent and infrequent elements in an artificial language. Furthermore, Gervain (2007) reported that adult speakers of head-final languages (e.g., Basque) chose a frequent-final segmentation preference of the artificial language more often than speakers of head-initial languages (e.g., French). The relative order of frequent (i.e., functors) and infrequent elements (i.e., content words) in the native language appeared thus to influence adult speakers' segmentation preferences of the artificial language.

The present chapter aims to investigate whether adult bilingual speakers of a head-final language (i.e., Basque) and a head-initial language (i.e., Spanish) deploy only one segmentation strategy, determined by their L1, or deploy the two segmentation strategies characteristic of the languages they command. To that end, in this chapter I present the results from an experiment conducted with early, highly proficient Basque/Spanish bilingual speakers using the artificial language created and used in Gervain (2007). In order to examine whether speakers of previously untested languages segment the artificial language as predicted by the frequency-based bootstrapping hypothesis, I test monolingual speakers of English and Spanish (both VO languages). Additionally, I also include in the experiment a group of French monolinguals (VO), a type of population that was also tested by Gervain (2007).

As presented in Chapter 2, section 2.1.2, functors are highly frequent elements in natural languages that typically occur at the edges of phrases (Morgan, Shi and Allopenna, 1996). Corpus analyses have shown that the frequency distribution of functors and the order of frequent and infrequent words at utterance edges reliably cue the basic word order of the languages (Gervain, 2007; Gervain et al., submitted), as predicted if functors head phrases and headedness has a tendency to have the same direction (first or last) in grammars (yielding head initial or head final grammars). Thus, Gervain (2007) and Gervain et al. (submitted) examined the distribution of functor and content words at utterance edges in an analysis of infant-directed speech corpora in Japanese, Basque[7] (OV, head-final languages), Italian and French (VO, head-initial languages). The analysis revealed that the two OV languages had more functor-final utterances than functor-initial utterances, whereas the opposite distribution was found in the VO languages. Within OV languages, Basque had more functor-final utterances than Japanese, whereas the two VO languages did not differ with respect to each other.

Gervain (2007) examined whether infants are indeed sensitive to the relative distribution of functors and content words and link these representations to the word order of their native language. To that end, Gervain (2007) tested the word order preferences of 8-month-old Japanese (OV, head-final) and Italian (VO, head-initial) infants in an artificial language learning experiment using the Headturn Preference Procedure (HPP). A four-syllable long basic unit was designed, with the structure *aXbY*. Categories *a* and *b* consisted of one CV token each (*a = fi; b = ge*); categories *X* and *Y* consisted of 9 CV tokens each (e.g., *X = lu, pe, pa*...; *Y = mu, li, bi*…). A 3 minute and 53 second long familiarization stream was created through concatenation of four-syllable long units. Hence, the tokens in categories *a* and *b* occurred 9 times more frequently (frequent categories) in the stream than the tokens from categories *X* and *Y* (infrequent categories). The frequency distribution of frequent and infrequent categories thus mirrored the frequency distribution in natural languages of functors and content words respectively.

Intensity was gradually increased in the initial 15 seconds of the familiarization stream and decreased in the final 15 seconds, which resulted in an ambiguous stream with two possible segmentations, depending on the ordering of the frequent and infrequent tokens: (a) a frequent-final segmentation, mirroring the frequency distribution and order of functors and content words found in OV languages, that is, ...*a [Xb] [Yc] Z*… or (b) a frequent-initial segmentation, mirroring the distribution and order of functors and content words found in VO languages, i.e., ...*[aX] [bY] c*.... (see Table 1 for a representation of the two possible segmentations of the familiarization in

[7] With the exception of Basque, where the corpus analysis was conducted on an extract of the Basque translation of *Harry Potter and the Philosopher's Stone.*

Gervain, 2007.) The familiarization stream was synthesized using a French voice (fr4, French female, MBROLA software) to a monotonous f0 of 200 Hz and a constant phoneme duration of 120 msec. Test stimuli consisted of 8 four-syllable long items, four with a frequent-initial structure (e.g., *aXbY: geMUfiDE*), and four with a frequent-final structure (e.g., *XbYa: DUgeRIfi*). After familiarization to the stream, infants were presented with 8 test trials, each trial consisting of 15 repetitions of a single test item.

Table 1. Familiarization stream (from Gervain, 2007)

Structure	…aXbYaXbYaXbYaXbYaXbYaXbY…
Stream	…**fi**lu**ge**mu**fi**pe**ge**li**fi**du**ge**ro**fi**pa**ge**bi…
Ambiguity:	
Frequent-initial	…**[**_**fi**lu**ge**mu_**]fi**pe**ge**li**[**_**fi**du**ge**ro_**]**_**fi**pa**ge**bi_…
	OR
Frequent-final	…**fi[**_lu**ge**mu**fi**_**]**pe**ge**li**fi**[_du**ge**ro**fi**_**]**pa**ge**bi…

Measurement of the looking times showed that the group of Japanese infants looked significantly longer to the frequent-final test items than to the frequent-initial items, whereas the group of Italian 8-month-olds looked longer to the frequent-initial items than to the frequent-final items. This opposite preference in the two groups of infants differed significantly. The 8-month-old infants were hence able to track the relative frequency of the elements present in the input. Furthermore, infants segmented the artificial language according to the relative order of frequent and infrequent words characteristic of their native languages. Hence, infants exposed to languages with opposite head directionality/word orders (Japanese: head-final, OV; Italian: head-initial, VO) have opposite preferences for the segmentation of a continuous string based on the relative order of frequent and infrequent elements. Gervain (2007) argues that this result suggests that infants might build an initial prelexical representation of the basic word order of phrases based on the frequency and relative order of functors and content words of the surrounding linguistic input.

Gervain (2007) examined whether adults also rely on the frequency and relative ordering of functional and content words to segment novel input, i.e., if the frequency-based bootstrapping mechanism is available in adulthood. To that end, Gervain (2007) conducted an artificial language learning experiment with adult speakers of several languages. Participants were native speakers of two OV, head-final languages (Basque and Japanese), two VO, head-initial languages (Italian and French) and one language with mixed order but predominantly head-final (Hungarian).

The basic unit consisted of a six-syllable long unit with the structure *aXbYcZ*, instead of the four-syllable long unit used in the study with infants. Frequent categories *a*, *b* and *c* consisted of one CV token each (*a = fi; b = nu; c=ge*); infrequent categories *X*, *Y* and *Z* consisted of 9 CV tokens each (e.g., *X = ru, pe; Y = fe, ta, Z = mu*, *ri*). This basic unit was concatenated resulting in a 17 minute and 30 second long familiarization stream. Tokens in the frequent categories occurred hence 9 times more frequently in the stream than tokens from the infrequent categories. Ambiguity of the stream was achieved through gradual increase and decrease of the initial and final 15 seconds. Familiarization stream and test items were synthesized using the es1 voice (Spanish male) of the MBROLA database (Dutoit, 1997) to a monotonous f0 of 100 Hz and a phoneme duration of 120 msec. Test stimuli consisted of 36 test pairs of six-syllable long sequences. Each pair consisted of an item with a frequent-initial structure (e.g., *aXbYcZ*: *geKUfiPAnuFE*), the other member having a frequent-final structure (*XbYcZa*, e.g., *TOnuKOgeRIfi*). After familiarization, participants listened to the test pairs and had to choose which member of each pair they thought sounded more like a possible "sentence" of the language heard during familiarization.

The groups of Japanese and Basque speakers, i.e., the speakers of OV, head-final languages, and the group of Hungarian speakers, i.e., speakers of a predominantly head-final language, showed a preference for the frequent-final items[8] (mean in Basque: 25.75[9] out of 36 trials; Hungarian: 23.21 out of 36; Japanese: 21.75 out of 36). The group of Basque participants were also bilingual speakers of Spanish, a head-initial language. Interestingly, this group still showed the strongest head-final preference. The groups of Italian and French speakers, i.e., the speakers of VO, head-initial languages, chose a frequent-initial segmentation more often than a frequent-final segmentation, though this preference did not differ from chance (mean in Italian: 17.05 out of 36 trials; French: 16.95 out of 36). Hence, speakers of languages with opposite head

[8] Separate means for each group are not reported by Gervain (2007), except for the mean of frequent-final responses in Hungarian. The individual means are taken from Gervain et al. (submitted). Instead, Gervain (2007) reports the mean of frequent-final responses in the two OV languages (22.81 out of 36 trials) and the mean of frequent-final responses in the two VO languages (17.2 out of 36 trials).

[9] Percentages of frequent-final preference are not reported in Gervain (2007), but means of frequent-final responses out of the 36 trials.

directionality/word orders showed different segmentation preferences for the ambiguous artificial stream.

Gervain (2007) argued that the absence of a clear frequent-initial preference in the speakers of VO languages might result from the fact that both Italian and French allow suffixation (inflectional and derivational). This frequency distribution in the morphology of the native languages might be influencing the participants' segmentation preference. Japanese, Basque and Hungarian are also characterized by a suffixing morphology. Being OV languages, this distribution, if exerting an influence, would strengthen the frequent-final preference. Indeed, the stronger frequent-final segmentation preference found in the group of Basque speakers as compared to the group of Japanese speakers tallies with the greater amount of frequent-final utterances reported by Gervain et al. (submitted) in the above mentioned corpus analysis in Basque. The different segmentation preference in OV and VO languages shows that adult speakers are sensitive to the relative frequency and distribution of functors in their native language, which are used as anchoring points to parse the artificial language.

3.2 Monolingual and bilingual segmentation preferences

As discussed in section 3.1, Gervain (2007) found that speakers of VO and OV languages segmented an artificial language according to the relative order of functor and content words in their native languages, following a frequency-based learning mechanism. The Basque speakers tested by Gervain (2007) were bilingual Basque/Spanish speakers, with Basque as L1 and Spanish as L2. Basque is an OV language and Spanish is a VO language. This group of bilinguals received the instructions and was addressed in their L1, i.e., Basque. The frequent-final preference found by Gervain (2007) in this group opens the ground for inquiring whether OV-VO bilinguals deploy only one segmentation strategy, determined by their L1, or whether they can deploy both. To answer this question, I explore the segmentation preferences of Basque/Spanish bilinguals, in order to determine whether OV-VO bilinguals (a) can use the segmentation strategies of the two languages they command and use frequently, or (b) prefer the segmentation strategy characteristic of their L1. To that end, I examine four groups of Basque/Spanish bilinguals, which differ in their L1 and L2

(L1Basque/L2Spanish or L1Spanish/L2Basque) and in the language in which they were addressed and received the instructions of the experiment (i.e., context language) (Basque or Spanish). Research on lexical access has provided ample evidence on the simultaneous activation of the bilinguals' two languages (Colomé, 2001; Costa, Caramazza and Sebastián-Gallés, 2000).

Additionally the present experiment aims to further examine whether speakers of VO and OV languages make use of the statistical cue provided by the frequency distribution of frequent and infrequent elements, as proposed by the frequency-based bootstrapping hypothesis, with a replica of Gervain's (2007) experiment with speakers of four languages, i.e., two languages that were previously tested in Gervain's study (Basque: OV, French: VO) and two languages that have been previously untested (Spanish: VO, English: VO).

If participants segment the artificial language according to the relative order of functors and content words within phrases in their native language, a different segmentation preference is predicted between Basque (OV) native participants and Spanish, French and English (all VO) native participants. Furthermore, if, as proposed by Gervain (2007), the distribution of morphological affixation of the languages exerts an influence in the segmentation preferences, English participants might display a greater frequent-initial segmentation preference than French and Spanish participants; though suffixation is present in the derivational morphology, English inflectional suffixation is virtually nonexistent.

3.2.1 Methodology

3.2.1.1 Participants

191 undergraduate students (128 females) from the University of the Basque Country (UPV/EHU), Reading University, and the Institut Universitaire de Technologie de Bayonne et du Pays Basque (IUT), with a mean age of 22.31 years of age (range 18 to 41) took part in this experiment. Participants were monolingual speakers of Spanish, English, or French and highly proficient Basque/Spanish bilinguals. 29 Spanish monolinguals (19 females) took part in the experiment, with a mean age of 25.93 years of age (range 18 to 41). The group of English monolinguals consisted of 26 participants

(17 females), with a mean age of 22.73 years of age (age range 18 to 40). 25 French monolinguals (17 females) took part in the experiment, with a mean age of 18.76 years of age (age range 18 to 20).

The Basque/Spanish bilingual participants were sorted into four groups, depending on (a) their L1, Spanish or Basque, and (b) the context language, i.e., the language in which they were addressed and received the instructions, Basque or Spanish. Thus, 20 L1Basque/L2Spanish bilinguals (16 females, mean age 19.1, age range 18 to 22) took part in the experiment, conducted in Basque and 28 L1Basque/L2Spanish bilinguals (18 females, mean age 23.29, age range 18 to 34) took part in the experiment, conducted in Spanish. Similarly, 32 L1Spanish/L2Basque bilinguals (21 females, mean age 21.84, age range 18 to 33) took part in the experiment conducted in Basque and 32 L1Spanish/L2Basque participants (20 females, mean age 23.1, age range 18 to 35) were addressed in Spanish.

Data concerning the linguistic background of the participants were collected by means of a questionnaire[10] that inquired about the acquisition and use of the languages by the participants. This questionnaire was developed by members of The Bilingual Mind at the University of the Basque Country (UPV/EHU).

Both groups of bilinguals (L1Spanish and L1Basque) in this study reported in the questionnaires having acquired their L2 at around 3 to 5 years of age. L1 and dominant Basque speakers usually have a highly proficient knowledge of Spanish. In the Basque Country both Spanish and Basque are spoken, though only 32% of the population is a highly proficient bilingual (V Sociolinguistic Survey 2011, Basque Government[11]). Since Spanish is legally required and prevailing in the cultural media (e.g., television, cinema), it is nearly impossible to find strict Basque monolinguals beyond infancy and childhood.

All L1Spanish/L2 Basque participants (a) had completed all their studies in Basque or were conducting studies in Basque at the UPV/EHU, (b) possessed a degree in Basque studies, or (c) had the EGA (Basque Language Certificate, C1 level in the Common European Framework), measures that certify their high proficiency in Basque. All participants were naïve to the purpose of the experiment and received a 4€ compensation for their participation.

---

[10] Questionnaire available in Appendix 1.

[11] http://www.euskara.euskadi.net/r59-738/es/contenidos/noticia/eae_inkestaren_laburpena_2011/es_argitalp/adjuntos/CAV_Resumen_de_la_encuesta_2011.pdf

3.2.1.2 Materials

The ambiguous artificial language used in this experiment was the one originally designed and used by Gervain (2007), which has been described in section 3.1 in the present chapter. Recall that this language contained two types of categories, characterized by a different frequency of occurrence in the artificial language, mimicking the relative frequency of functional and content words in natural languages:

(a) Frequent categories: consisted of three categories (*a*, *b*, *c*), where each category contained only one CV syllable token (*a* = *fi*; *b* = *nu*; *c* = *ge*).

(b) Infrequent categories: consisted of three categories (*X*, *Y*, *Z*), where each category contained 9 CV syllable tokens, as shown in Table 2.

Frequent categories were nine times more frequent than infrequent categories. (see Tables 2 and 3)

The language comprised a total of 3 frequent tokens (i.e., three categories, one token each) and 27 infrequent tokens (i.e., three categories, nine tokens each). The language was constructed concatenating hexasyllabic units with the structure *aXbYcZ*, i.e., characterized by a strict alternation of frequent and infrequent categories.

Table 2. Structure of the hexasyllabic units: aXbYcZ

	Frequent categories	**Infrequent categories**
Categories	a, b, c	X, Y, Z
Number of tokens	3	27
Number of tokens per category	a = 1; b = 1; c = 1	X = 9; Y = 9; Z = 9
Syllabic structure of the categories and tokens	CV	CV
Relative frequency of the categories	Frequent categories are 9-times more frequent than infrequent categories.	

Table 3. Experiment A. Lexicon of the CV syllables (from Gervain, 2007)

a	**X**	**b**	**Y**	**c**	**Z**
	ru		fe		mu
	pe		ta		ri
	du		pi		ku
	ba		be		bo
fi	fo	nu	bu	ge	bi
	de		ko		do
	pa		mo		ka
	ra		po		na
	to		pu		ro

This basic unit was concatenated 540 times, resulting in a 17 minute and 30 second long stream. The intensity of the initial 15 seconds of the stream was gradually increased and the final 15 seconds gradually decreased, which rendered the stream ambiguous, allowing two possible segmenting schemas: (a) a frequent-initial segmentation (*aXbYcZaXbYcZ…*) that mirrored the relative order of frequent (functors) and infrequent (content words) elements in VO languages, or (b) a frequent-final segmentation (*XbYcZaXbYcZa…*), which mirrored the relative order of frequent and infrequent elements in OV languages. (see Table 4)

Table 4. Experiment A. Familiarization (from Gervain, 2007)

Structure	…aXbYcZaXbYcZaXbYcZaXbYcZaXbYcZaXbYcZ…
Stream	…filunufegemufipenutagelifidunupigerofipanumogebi…
Ambiguity:	
Frequent-initial	…**[*fi*ilu*nu*fe*ge*mu]fi**pe**nu**ta**ge**li**[*fi*du*nu*pi*ge*ro]fi**pa**nu**mo**ge**…
	OR
Frequent-final	…**fi[**lu*nu*fe*ge*mu*fi***]**pe**nu**ta**ge**li**fi[**du*nu*pi*ge*ro*fi***]**pa**nu**mo**ge**…

Test stimuli consisted of 36 six-syllable long items, 18 with a frequent-initial structure (i.e., started with a frequent category), 18 with a frequent-final order (i.e., started with an infrequent category). In Gervain (2007) stimuli were synthesized with a Spanish voice (es1 of the MBROLA database). In the present experiment, however, both the familiarization and the test items were synthesized with the de6 (German male) voice of the MBROLA database (Dutoit, 1997), i.e., with a voice in an unknown language to all participants. The phonemes had a constant duration of 120 msec and a monotonous f0 of 100 Hz.

3.2.1.3 Procedure

The Spanish monolingual group and all four bilingual groups were tested in the psycholinguistics laboratory at the University of the Basque Country, UPV/EHU. The experiments with English monolinguals were conducted in the Centre for Integrative Neurosciences and Neurodynamics (CINN) directed by Professor Douglas Saddy, at Reading University (UK). The experiments with French monolinguals were conducted in the Institut Universitaire de Techonologie de Bayonne et du Pays Basque (IUT) of the Université de Pau et des Pays de l'Adour (France), with the help of Dr. Urtzi Etxeberria and Dr. Jasone Salaberria (IKER-CNRS), and María Arana and Murielle Fraty (UIT Bayonne). All participants were tested individually in a quiet room, except for the group of French monolinguals, who were tested collectively in two sessions, in the language laboratory located at UIT (Bayonne), containing individual booths equipped with computers and headsets. In all instances, the experimental input was displayed in a computer screen and participants were provided with a set of high quality headphones.

Participants read the instructions on the computer screen. They were told that they were about to listen to an unknown language for several minutes and, that, afterwards, they would listen to "questions" regarding the language they had heard. As noted above, in the experiments with the Basque/Spanish bilingual population, the language used during the instructions was counterbalanced. Half of the L1Basque/L2Spanish participants was addressed in Basque, their L1, whereas the remaining half of the participants was addressed in Spanish, their L2. Similarly, half of the L1Spanish/L2 Basque participants received their instructions in Spanish, their L1, whereas the remaining half received these in Basque, their L2.

Once the instructions were clear, a short training was carried out, in order to familiarize participants with the procedure of the experiment. They heard six pairs of monosyllables; one of the members of each pair was the syllable *so*. Participants' task was to identify this syllable in every pair and press one of two predefined keys in a keyboard, depending on whether the target syllable had been heard first, or second. Both positive and negative feedback were provided.

After the training session, participants found a new screen that reminded them of the instructions of the experiment. When ready, participants pressed the space-bar to start with the familiarization process, in which participants listened to the artificial language during 17 minutes and 30 seconds. During the familiarization the screen in the computer was blank.

After familiarization, a new message appeared on the computer screen that reminded participants of the task and asked them to press the space-bar to start with the test phase. Participants listened to 36 test trials. In each trial a pair of six-syllable long test sequences were presented, one sequence always being frequent-initial, the other frequent-final. The two sequences were separated by a 500 msec pause. Each six-syllable sequence appeared twice during the experiment, one time as the first member of a pair, and another time as the second member of a pair, but never in consecutive test trials. The order of the stimuli was randomized. The participants' task consisted on choosing the sequence in each pair that they thought belonged to the language they had heard before, in the familiarization phase. They had to express their choice by pressing the corresponding key in a keyboard. After the 36 experimental trials a last message displayed on the screen informed the participants of the end of the experiment. The session had a total duration of around 30 minutes.

3.2.2 Results

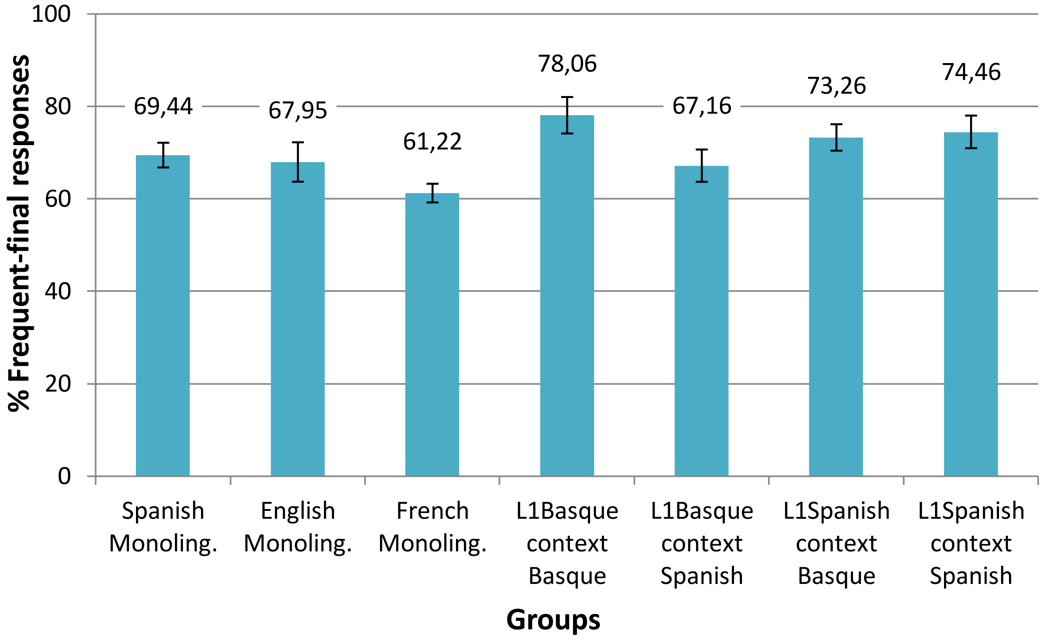
The program recorded the number of frequent-final responses per participant out of the 36 test trials. All groups showed a general preference for the frequent-final segmentation. The group of L1Basque bilinguals addressed in Basque showed the greatest frequent-final segmentation preference, whereas the group of French monolinguals chose the frequent-final segmentation least often. (see Table 5)

Table 5. Experiment A. Mean of frequent-final preference per group

Group	Frequent-final responses (out of 36)	Percentage of frequent-final responses
L1Basque context Basque	28.1	78.06%
L1Basque context Spanish	24.18	67.16%
L1Spanish context Basque	26.38	73.26%
L1Spanish context Spanish	26.81	74.46%
Spanish monolinguals	25	69.44%
English monolinguals	24.46	67.95%
French monolinguals	22.04	61.22%

Data analysis was carried out with the SPSS software. A Kolmogórov-Smirnov test of normality, conducted for each of the seven groups, revealed normal distribution of the dependent variable in all groups. A one-sample t-test was conducted for each group to observe if their segmentation preference differed significantly from chance. The mean score of the sample was compared to a test value of 18 (out of 36, a value of 18 would amount to chance). The t-tests showed that the segmentation preference in all groups was significantly above chance (all $p < .001$), i.e., a frequent-final segmentation preference was found in all groups, as shown in Graph 1.

Graph 1. Experiment A. Percentage of frequent-final responses per group.[12]



A one-way ANOVA with the factor *group* was carried out and a significant main effect was obtained (F(6.184) = 2.530, p = .022). Pair-wise comparison of the groups was conducted by means of separate independent sample t-tests. The results of the t-tests will be addressed separately in what follows.

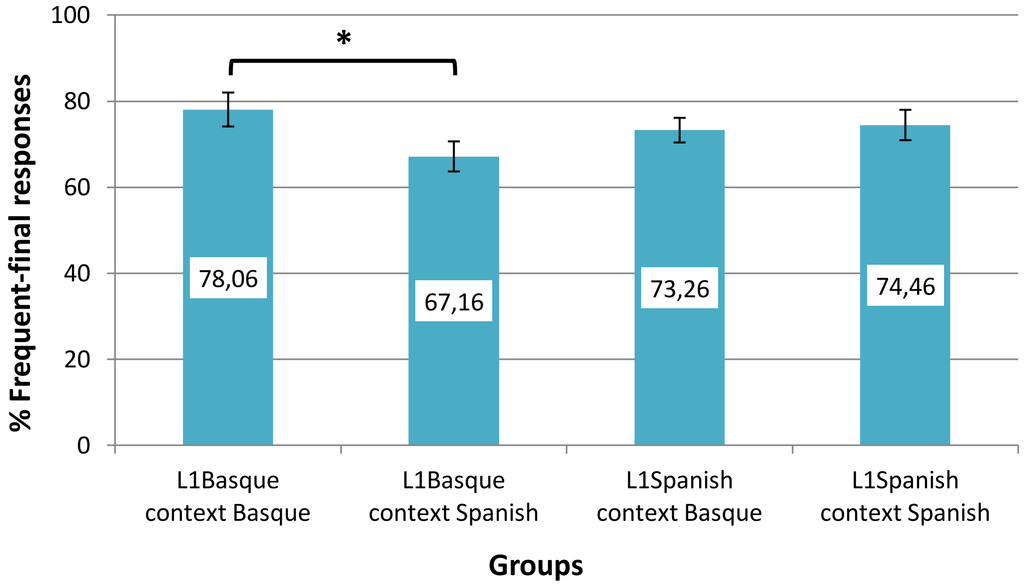
3.2.2.1 Context language

This experiment aimed to examine whether the context language (i.e., language of the instructions) influenced the bilinguals' segmentation preference. To that end, pair-wise comparisons were conducted between the groups of bilingual participants. The group of L1Basque context Basque bilinguals (M[13] = 28.1, SE = 1.42) chose a frequent-final segmentation significantly more often than the group L1Basque context Spanish bilinguals (M = 24.18, SE = 1.26) (t(46) = 2.055, p = .046). However, no significant preference was found between the two groups of L1Spanish bilinguals, i.e.,

[12] The error bars correspond henceforth to two standard errors.
[13] *M* refers to the mean of frequent-final responses. *SE* refers to the standard error.

L1Spanish context Basque (M = 26.38, SE = 1.03) and L1Spanish context Spanish (M = 26.81, SE = 1.27) ($t(61) = -.265$, $p = .792$), as shown in Graph 2.

Graph 2. Context languages' influence in segmentation preferences

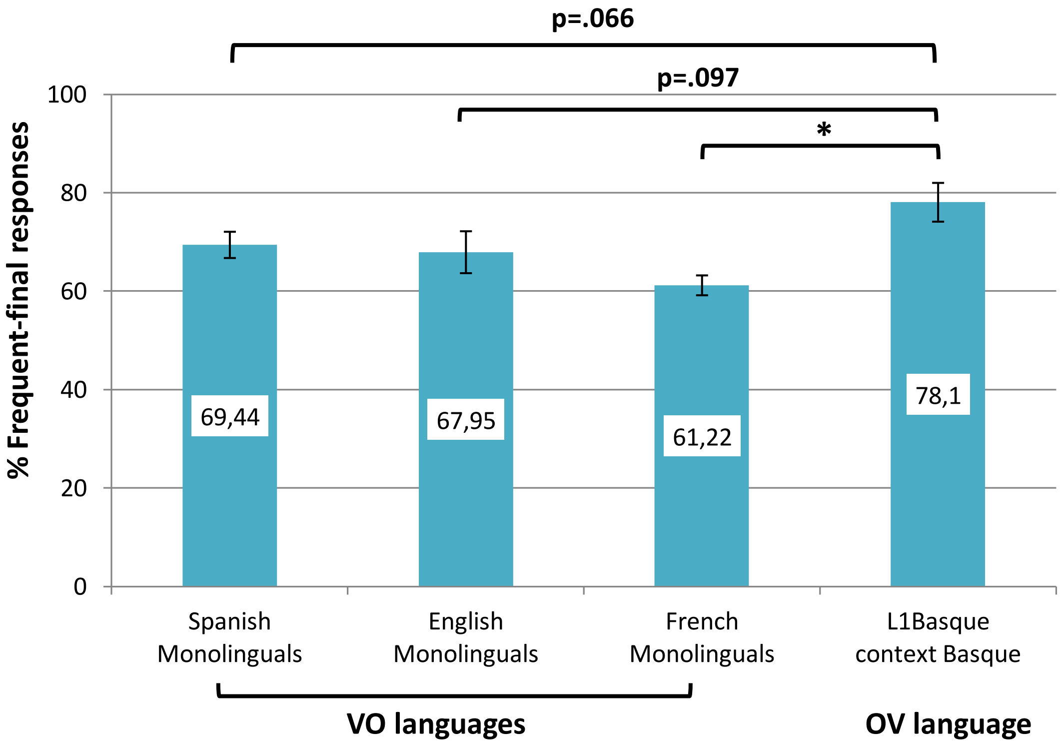


Hence, language of the context significantly influenced the segmentation preferences of the L1Basque bilinguals, but not the segmentation preferences of L1Spanish bilinguals.

3.2.2.2 OV vs. VO languages

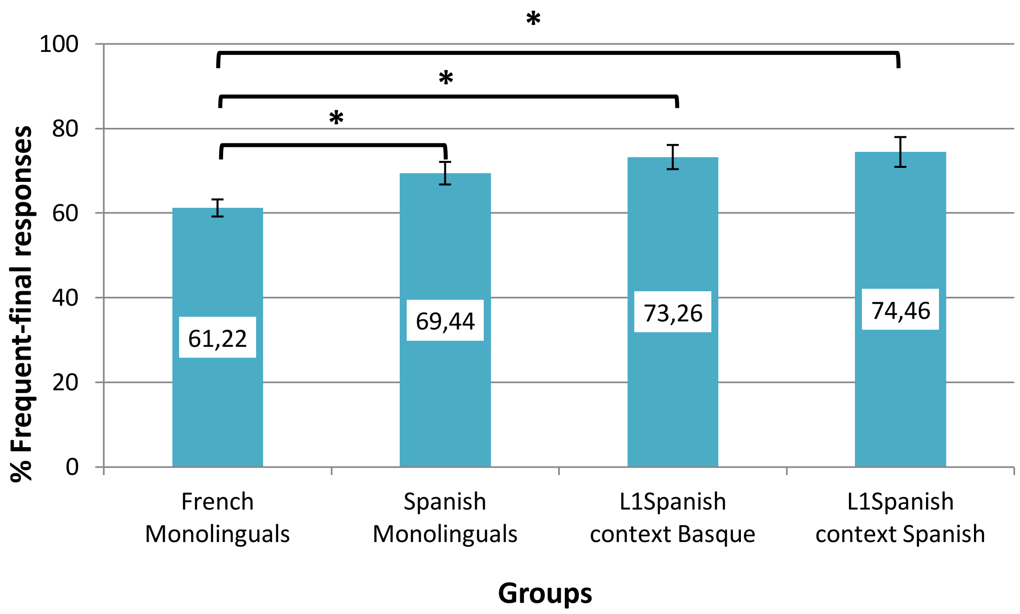
If participants segment the artificial language according to the relative order of functors and content words characteristic of their native languages, a difference between L1Basque (OV) and French, Spanish and English (VO) native participants would be expected, independently from other experimental conditions. Pair-wise comparison revealed differences in the segmentation preferences of VO and OV languages, as shown in Graph 3. Thus, the group of L1Basque context Basque (M = 28.1, SE = 1.42) chose the frequent-final segmentation significantly more often than the group of French monolinguals (M = 22.04, SE = .73) ($t(43) = 4.029$, $p < .001$). The group of L1Basque context Basque (M = 28.1, SE = 1.42) showed also a greater frequency of frequent-final responses than the groups of Spanish monolinguals (M = 25, SE = .96) and English monolinguals (M = 24.46, SE = 1.53), though these differences were only marginally significant (L1Basque context Basque vs. Spanish monolinguals: t(47) = -1.881, p = .066; L1Basque context Basque vs. English monolinguals: t(44) = 1.695, p = .097).

Graph 3. Segmentation preferences in VO vs. OV languages



Pair-wise comparisons revealed further significant differences in the segmentation preferences, as shown in Graph 4. Thus, the group of French monolinguals (M = 22.04, SE = .73), group that had shown the smallest frequent-final segmentation preference, chose a frequent-final segmentation significantly less than both groups of L1Spanish bilinguals, i.e., L1Spanish context Basque (M = 26.38, SE = 1.03) (t(55) = 3.256, p = .002) and L1Spanish context Spanish (M = 26.81, SE = 1.27) (t(54) = 3.056, p = .003), as well as the group of Spanish monolinguals (M = 25, SE = .96) (t(52) = 2.397, p = .020). The remaining pair-wise comparisons did not yield significant results.

Graph 4. Further significant differences in segmentation preference



3.3 Discussion

3.3.1 Context language

This experiment aimed to investigate whether bilingual speakers of Basque (OV) and Spanish (VO) (a) used the segmentation strategies of the two languages they command, or (b) preferred the segmentation strategy characteristic of their L1. The significant difference found within the group of L1Basque bilinguals revealed that the context language (that is, the language in which participants were addressed and received the instructions) significantly influenced their performance: the group given instructions in Basque produced a significantly greater number of frequent-final responses (78.06%) than the group given instructions in Spanish (67.16%).

No such difference was found in the responses of the L1Spanish bilinguals, i.e., language context did not modulate the L1Spanish bilinguals' segmentation preferences. A potential explanation for this difference in the performance of L1Basque and L1Spanish bilinguals might be found in differences in the proficiency and frequency of use of the two languages by the speakers of these two groups.

To that end, the linguistic background of the participants collected in questionnaires was examined. (see Appendix 1) As reported in the methodology, the questionnaires showed that acquisition of the L2 (Basque in L1Spanish bilinguals, Spanish in L1Basque bilinguals) took place at around 3 to 5 years of age in both groups of bilinguals, mainly at the start of schooling. Both L1Basque and L1Spanish bilingual participants received primary and secondary education in Basque.[14] Additionally, great care was taken to ensure that L1Spanish bilinguals had been raised in Spanish-speaking homes, whereas L1Basque bilinguals had been raised in Basque-speaking homes.

Besides these variables, the questionnaires inquired about the frequency of use of the two languages in other situations (outside from school and home), at three stages of development: during infancy (i.e., primary education), adolescence (i.e., secondary education) and adulthood (i.e., at the time of the experiment). Analysis of the questionnaires revealed that L1Basque bilinguals used Basque in the majority of situations, whereas the L1Spanish bilinguals generally used Spanish in their interactions. Thus, participants rated the frequency of use of their languages by choosing one of seven options. For the purposes of the analysis, a numerical value was given to each option. (see Table 6)

Table 6. List of options and correspondent numerical value

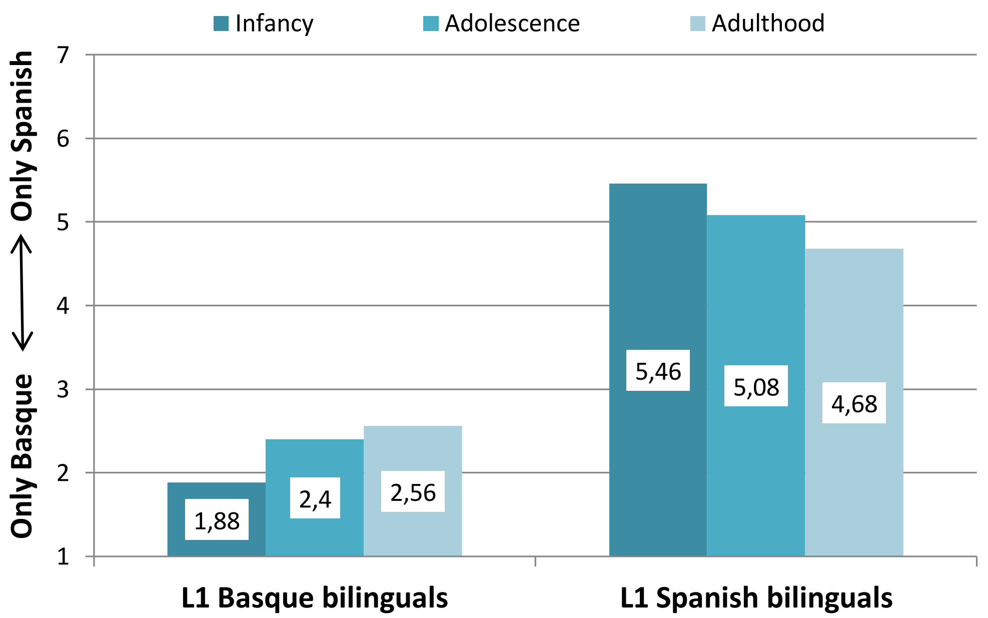
Options	Value
Only Basque.	1
Almost always Basque, rarely Spanish.	2
Mainly Basque, albeit using Spanish at least 25% of the time.	3
Basque and Spanish with equal frequency.	4
Mainly Spanish, albeit using Basque at least 25% of the time.	5
Almost always Spanish, rarely Basque.	6
Only Spanish.	7

Contingency table analyses revealed that, within the group of L1Basque bilinguals, the frequency of use of the two languages of the group given the instructions in Basque did not differ, in any of the three stages, from the frequency of use of the group given the instructions in Spanish.[15] Neither did the frequency use of the two

[14] Note however that the study of Spanish is a mandatory subject in the Spanish educational system.
[15] Statistical analysis is reported in detail in Appendix 4.

groups of L1Spanish bilinguals. Hence, these subgroups were collapsed into two main groups, that is, L1Basque bilinguals and L1Spanish bilinguals. Significant differences were found between L1Basque and L1Spanish bilinguals' frequency of use of their two languages, in all three points of development. The group of L1Basque bilinguals reported predominant use of Basque in situations other than at school or at home, whereas the group of L1Spanish bilinguals showed the opposite pattern, a predominant use of Spanish in their interactions (see Graph 5).

Graph 5. Frequency of use of Basque and Spanish



The vertical axis in Graph 5 depicts the response options given to the participants, presented as a continuum ranging from 1 to 7, in which 1 stands for exclusive use of Basque in situations other than at school/home and 7 stands for exclusive use of Spanish.

Hence, this opposite patterns of language predominance found in the groups of L1Basque and L1Spanish bilinguals, added to the particular sociolinguistic situation of the Basque Country, could account for differences in the two groups' proficiency of the two languages.

Basque and Spanish coexist in the Basque Country in a diglossic situation (Etxebarria, 2004). According to the latest sociolinguistic survey[16] (V Sociolinguistic

[16] http://www.euskara.euskadi.net/r59-738/es/contenidos/noticia/eae_inkestaren_laburpena_2011/es_argitalp/adjuntos/CAV_Resumen_de_la_encuesta_2011.pdf

Survey 2011, Basque Government) only 32% of the population over 16 years in the Basque Country is highly proficient in both languages, whereas 50.8% are monolingual Spanish speakers. Only 17.5% of the population has Basque as their first language. Furthermore, only 20% of the population of the Basque Country report using Basque with equal or superior frequency than Spanish, whereas 71.1% of the population exclusively use Spanish. The population between 16–34 years of age (i.e., range similar to the participants' age range in the present experiment), show the greatest proportion of bilingualism. Thus, 59.7% of the 16- to 24-year-olds and 44.5% of the 25- to 34-year-olds are Basque/Spanish bilinguals.

Regarding the geographical distribution of the bilingual population, half of the Basque/Spanish bilingual population is concentrated in Gipuzkoa (49.9%), one of the three provinces that form the Basque Country. 25 of the 48 L1Basque bilingual participants in the experiment came from this province, mirroring this geographical distribution. However, only 8 of the 62 L1Spanish bilingual participants came from Gipuzkoa. Also, the bilingual population in the Basque Country is typically concentrated in areas with low density populations, rather than on province capitals. Again, the background of the L1Basque participants, unlike the background of L1Spanish bilinguals, tallied with the areas of concentration of Basque/Spanish bilinguals. Only 6 of the 48 L1Basque participants in the experiment came from a province capital, as opposed to 32 of the 62 L1Spanish bilingual participants, of which 17 came from Vitoria, the capital with the lowest percentage of highly proficient bilinguals in the Basque Country (15.5%).

As mentioned in the methodology, all L1Spanish bilingual participants had either completed all their studies in Basque or were conducting studies in Basque at the university, possessed a degree in Basque studies, or had the EGA (Basque Language Certificate, C1 level in the Common European Framework), measures that ensured proficiency in their L2, Basque. On the other hand, the L1Spanish bilinguals' use of Basque is quite limited, due to the prevalence of Spanish in the areas of origin of these bilinguals, as shown above. It is precisely this prevalence which guarantees the high proficiency of Spanish in the group of L1Basque bilinguals. Even in the areas were 50–80% of the population are Basque/Spanish bilinguals, over one third of the population reports using only Spanish in their interactions (37.2%).

Research on lexical access has revealed differences in language switching cost between highly proficient and less proficient bilinguals (Costa and Santesteban, 2004;

Meuter and Allport, 1999). As mentioned in the introduction, during lexical access the two languages of the bilingual are simultaneously activated (Colomé, 2001; Costa, Caramazza and Sebastián-Gallés, 2000). Thus, Colomé (2001) conducted a series of phoneme monitoring tasks with highly proficient Catalan/Spanish bilinguals in which participants were presented with pictures (e.g., a table) and had to decide whether a target phoneme (e.g., /*t*/) was present in the Catalan word designating the picture (e.g., *taula*, Catalan for *table*). Bilinguals rejected the phonemes that appeared in the Spanish word designating the picture (e.g., /*m*/ in *mesa*) slower than phonemes that did not occur in the Spanish and Catalan words (e.g., /*f*/), revealing hence activation of the bilinguals' two languages.

Studies with language switching tasks have revealed asymmetrical switching costs between the L1 and the L2 (Meuter and Allport, 1999). Thus, in a language switching naming task with bilingual participants from various backgrounds, Meuter and Allport (1999) found greater switching cost, i.e., longer reaction times (RTs), in the switch from a trial in the L2 to a trial in the L1, than in the switch from a trial from the L1 to the L2. Meuter and Allport (1999) argued that this asymmetrical cost resulted from the greater amount of inhibition required by the L1 as compared to the L2, in line with Green's Inhibitory Control model (1998). According to this model, activation of the words in the non-target language is inhibited during lexical access. The amount of inhibition required depends on the degree of activation of the lexical items determined by the bilingual's proficiency in the non-target language. Suppression of the activated lexical items in the L1 requires thus more inhibition than suppression of the less active L2 lexical items.

Costa and Santesteban (2004) replicated this asymmetrical language switching cost in a naming experiment with L1Spanish/L2 learners of Catalan and L1Korean/L2 learners of Spanish. However, Costa and Santesteban (2004) found a symmetrical language switching cost in two subsequent experiments with highly proficient speakers of L1Spanish/L2Catalan. Costa, Santesteban and Ivanova (2006) replicated this symmetrical switching cost in a naming task with L1Spanish/L2Basque and L1Spanish/L2English highly proficient speakers. Furthermore, Costa et al. (2006) found similar symmetrical switching costs between highly proficient bilinguals' L2 and their L3, language in which participants were less proficient, and between their L1 and L3 (i.e., L1Spanish/L2Catalan/L3English). These results suggest that achieving high proficiency in any of the bilinguals' languages entails processing differences, as compared to bilinguals who are low proficient in their non-native languages.

In order to explain the language context effect obtained in the group of L1Basque/L2Spanish bilinguals in the present experiment, it could be argued that the diglossic situation of the Basque Country demands frequent changes from Basque to Spanish in the group of L1Basque bilinguals, as opposed to the more limited use of Basque in the group of L1Spanish bilinguals, as shown by the analyses of the questionnaires. The more balanced use of Basque and Spanish in the group of L1Basque bilinguals might entail greater proficiency in the two languages, leading to a different processing of the artificial language.

There is another possible explanation to the absence of a context effect in L1Spanish bilinguals. It could be that frequent-initial ordering of frequent and infrequent elements is the default, or unmarked order, whereas frequent-final ordering is the marked order. Kayne (1994) proposes that natural languages have a universal underlying order in which specifiers precede heads, and, crucially, heads precede complements. Head-final languages would thus result from the movement of the complement to a specifier position. Chomsky (1995) further elaborates this hypothesis, proposing that the unmarked specifier-head-complement order proposed by Kayne (1994) corresponds to the SVO order in natural languages.

Langus and Nespor (2010) provide supporting evidence to this proposal in a speech comprehension task with speakers of an OV language (i.e., Turkish) and speakers of a VO language (i.e., Italian). Participants listened to strings of words in their native language, artificially synthesized and rendered prosodically flat. Stimuli were arranged into sequences that instantiated the six possible orderings of subjects, objects and verbs (i.e., SOV, VOS, VSO, OSV, OVS and SVO). After each sequence, participants had to choose which of two displayed vignettes matched the previously heard sequence. Crucially, both groups of participants chose the correct vignette faster after listening to sequences with VO order (collapsing the results from sequences in the three word orders in which O follows V, i.e., SVO, VOS, VSO) than after sequences with OV order (i.e., OSV, OVS and SVO). Hence, VO sequences showed a processing advantage.

Following this hypothesis, it could be argued that the performance differences between L1Basque and L1Spanish bilinguals might result from the fact that L1Basque bilinguals, native speakers of an OV language, have acquired the marked ordering of heads and complements (i.e., head-final). They might hence be able to implement the two strategies characteristic of their two languages, i.e., the marked strategy characteristic of their L1 (head-final) and the unmarked strategy that characterizes their L2, Spanish (head-initial). L1Spanish bilinguals, on the other hand, might not have acquired the marked strategy of the L2, and be therefore able to deploy only the segmentation strategy of their L1.

3.3.2 Segmentation preferences on VO vs. OV languages

The second aim of the present experiment was to further explore the statistically based segmentation strategies proposed by the frequency-based bootstrapping hypothesis (Gervain, 2007), with two previously untested languages, namely, Spanish and English, as well as two of the languages tested by Gervain (2007), i.e., Basque and French. According to this hypothesis, the speakers build representations of the relative frequency distribution and order of functors (i.e., frequent words) and content words (i.e., infrequent words) characteristic of their native languages and make use of these representations to segment new input, such as an ambiguous artificial language. Given that the languages' order of frequent and infrequent words correlates with the order of verb and object, differences in the segmentation preferences are expected in speakers of VO and OV languages, i.e., frequent-initial preference and frequent-final segmentation respectively.

As predicted by the frequency-based bootstrapping hypothesis, the group of L1Basque bilinguals, native speakers of an OV language, showed a marked preference for frequent-final segmentation of the ambiguous language, significantly differing from the segmentation preference of the group of French monolinguals (VO language) and hence replicating Gervain (2007). The speakers of the previously untested VO languages, i.e., Spanish and English, showed a tendency in the direction predicted, i.e., both groups chose the frequent-final segmentation less often than the group of L1Basque. However, these differences were only marginally significant, probably due to the general frequent-final bias observed in all populations.

The group of French monolinguals tested in this experiment showed a general frequent-final preference, unlike the frequent-initial segmentation preference[17] reported in Gervain (2007). This bias towards a frequent-final segmentation found in all groups will be addressed later in the discussion.

The results show that there is a correlation between the frequency distribution and relative order of functional and content word categories in the adult speakers' languages and the segmentation preferences for frequent and infrequent syllables in the artificial syllabic stream. The frequent-final order characteristic in Basque (OV) might thus have led the L1Basque bilinguals to prefer a frequent-final segmentation of the ambiguous language, as opposed to the smaller frequent-final segmentation preference found in the speakers of frequent-initial languages (i.e., Spanish, English and French).

These results show that adult speakers are able to track the relative frequency of the elements in the artificial language and make use of "functor-like" elements (in terms of frequency) to divide the input into phrase-like units. This finding provides supporting evidence on the claim that the salient frequency of functors in natural languages might lead listeners to treat them as anchoring points to the syntactic structure of the language (Green, 1979; Gervain, 2007; Gervain et al., submitted; Valian and Coulson, 1988).

Gervain (2007) argued that the non-significant frequent-initial preference found in the groups of French and Italian monolinguals might proceed from the amount of suffixation allowed in the morphology of these two languages. Thus, the adult listeners' ability to track frequency distributions captures different types of frequency distributions present in the languages, e.g., in morphology. Suffixation entails a frequent-final ordering, which might hence diminish the general frequent-initial representation of the French and Italian monolingual speakers. Indeed, across natural languages a universal preference for suffixes over prefixes or infixes has been attested (Greenberg, 1963; Hawkins and Gilligan, 1988). Thus, OV languages are characterized by having mostly or exclusively suffixes, whereas VO languages typically have both prefixes and suffixes.

A processing account for this cross-linguistic preference for suffixation was proposed by Cutler, Hawkins and Gilligan (1985). According to this proposal, listeners convey computational priority to stems, which are hence placed at a salient position within the word, i.e., the beginning. Stems convey lexical-semantic information,

[17] Though not significantly different from chance.

whereas suffixes convey syntactic information. Initial processing of the stem would thus enable faster word recognition than would initial processing of the suffix.

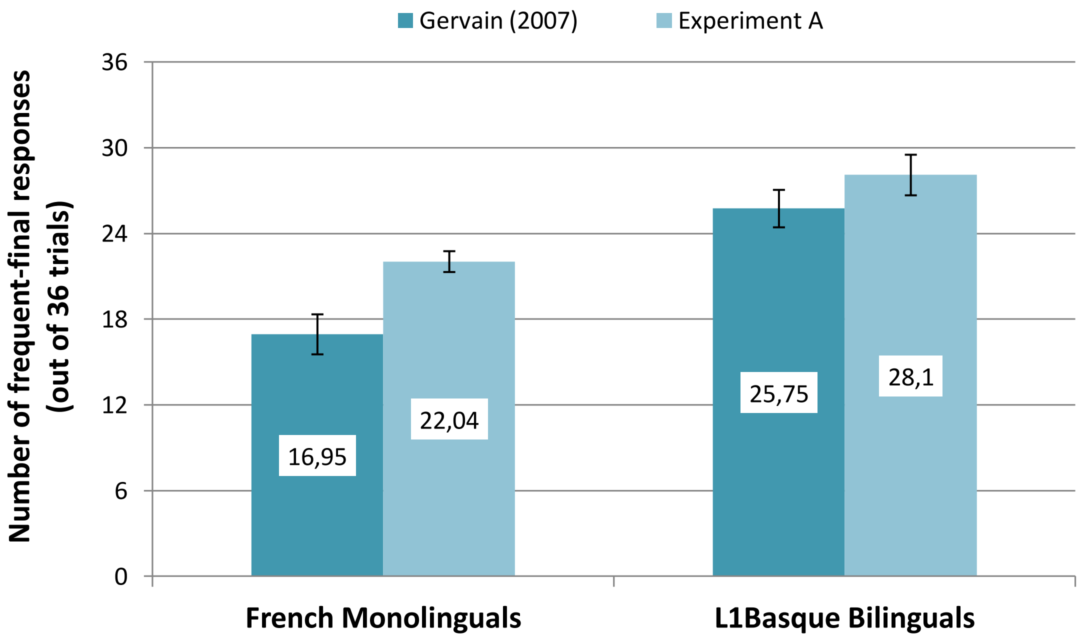
Gervain's (2007) proposal on the potential influence of suffixation in the segmentation preferences of speakers of VO languages is unlikely to account for all the differences found between the VO languages in the present experiment. As reported, the group of French monolinguals showed the lowest frequent-final preference of all groups tested (61.22%). The group of Spanish monolinguals, on the other hand, showed significantly greater frequent-final segmentation (67.95%). In order for suffixation to account for this difference, suffixation should be greater in Spanish than in French. Both languages have derivational suffixation (and prefixation) and exclusively suffixing inflections. However, some suffixes in French are only realized in liaison contexts, which might result in a greater amount of suffixes in Spanish than in French. Corpora analyses of the frequency distribution of suffixes and prefixes in these two languages could help clarify whether an asymmetry exists in the frequency distribution of suffixes and prefixes in these two languages.

The influence of suffixation seems unlikely to account for the preference obtained in the group of English monolinguals (67.95%), i.e., greater frequent-final segmentation than the group of French monolinguals (61.22%) though statistically not significant. Thus, even though the number of suffixes triples the number of prefixes in English (Fudge, 1984, as cited in St. Clair, Monaghan and Ramscar, 2009, lists 56 prefixes and 181 suffixes), inflectional morphology is extremely impoverished. The limited suffixation in English should thus entail a greater frequent-initial representation of the language by their speakers, which should result in a lower frequent-final preference than the found in the French monolinguals, contrary to the more marked frequent-final preference displayed by the group of English monolinguals. Again, a corpus analysis would shed light on the frequency distributions of suffixes and prefixes in English, though it seems unlikely that suffixation would exert an even greater influence on the English monolinguals than the more extensive use of suffixation in French in the group of French monolinguals.

3.3.3 General frequent-final bias

Gervain (2007) reported a frequent-final segmentation preference in OV and predominantly head-final languages, i.e., Japanese, Basque and Hungarian. The VO and head-initial languages tested (French and Italian), on the other hand, showed a frequent-initial preference, though not significantly different from chance. By contrast, a general, significant preference for frequent-final segmentation was found both in OV and VO languages in the present experiment. Indeed, the groups of L1Basque bilinguals context Basque and French monolinguals showed greater frequent-final segmentation preference in the present experiment than the two similar groups tested by Gervain (2007), as shown in Graph 6.

Graph 6. Comparison of the frequent-final preference in Gervain (2007) and Experiment A.[18]



Hence, participants in the present investigation showed a general bias towards frequent-final segmentation, though they significantly differed in the degree to which they preferred a frequent-final segmentation. Materials and procedure in the present experiment were the same as those used in Gervain (2007) with the exception of the

[18] As mentioned in section 3.1. of the present chapter, percentages of frequent-final preference are not reported in Gervain (2007) and Gervain et al. (submitted), but means of frequent-final responses out of the 36 trials.

voice used in synthesizing the stimuli. In Gervain (2007), stimuli were synthesized with a Spanish-based voice (es1 voice, Spanish male, MBROLA software), whereas the current stimuli were synthesized with a German-based voice (de6, German male, MBROLA software), foreign language to all participants. This change in the acoustic properties of the stimuli appears to have induced a general shift towards a frequent-final segmentation.

As described in Chapter 2, section 2.3, acoustic-phonetic properties of the input play an important role in adult's perception of speech, and particularly in word segmentation. In the last decades, a great deal of research has uncovered a variety of cues used in segmentation, revealing an array of phonological and statistical cues available to both adults and infants: transitional probabilities (TPs) (Saffran, Aslin and Newport, 1996; Saffran, Newport and Aslin, 1996), rhythm (Cutler, 2004; Nazzi et al., 2006), phonotactic constraints (Friederici and Wessels, 1993; Massaro and Cohen, 1983), allophonic variation (Jusczyk, Hohne and Bauman, 1999; Smith and Hawkins, 2000), or coarticulation (Johnson and Jusczyk, 2001; Mattys, White and Melhorn, 2005). Successful word segmentation depends on the listener's ability to integrate the available cues (Christiansen, Allen and Seidenberg, 1999).

As revealed by numerous studies that pitted these available cues against each other (see Chapter 2, section 2.3), listeners weigh some of the cues more heavily than others (Fernandes, Ventura and Kolinsky, 2007; Johnson and Jusczyk, 2001; Mattys, White and Melhorn, 2005; McQueen, 1998; Morgan and Saffran, 1995). Mattys, White and Melhorn (2005) propose that segmentation cues are hierarchically organized into a three-tiered structure, depending on the relative weight given to the different cues by the listeners, and that this hierarchy changes throughout development. Whenever available, adults rely primarily on lexical information in word segmentation. In the absence of lexical information, adults make use of segmental cues like phonotactics, allophonic variation and coarticulation. Rhythmic and statistical cues have been proposed to be used by adults only as a last resort, when the signal contains poor segmental information (e.g., in a noise degraded signal) (Fernandes, Ventura and Kolinsky, 2007; Mattys, White and Melhorn, 2005). On the other hand, rhythmic and statistical cues are the first available cues to infants, as shown in Chapter 2, due to the lack of lexical information and knowledge of segmental cues. Acquisition of segmental cues (e.g., phonotactics, allophonic variation, coarticulation) takes place during the second half of the first year of life, time at which perceptual reorganization is known to take place (Bosch and

Sebastián-Gallés, 2003; Polka and Werker, 1994; Werker and Lalonde, 1988; Werker and Tees, 1984).

The artificial language in the present experiment provided no lexical information that might have assisted adult participants in segmentation. The prevalence of segmental cues over statistical cues found in adult speech segmentation tallies with the differences in segmentation preferences found between Gervain (2007) and the current experiment, differences induced by the acoustic-phonetic properties of the stimuli, as I will argue in what follows.

Cross-linguistic differences in the realization of stops might provide an explanation for the frequent-final bias found in the present study. German, Spanish, Basque and French share voiced and voiceless stops (/*b*/, /*d*/, /*g*/, /*p*/, /*t*/, /*k*/), but stops in German are typically produced with a different Voice Onset Time (VOT) than stops in Spanish, Basque and French (Braunschweiler, 1997; Caramazza and Yeni-Komshian, 1974; Castañeda, 1986; Hualde, 2003). The VOT of a plosive refers to the period between the explosion of the stop and the beginning of the vocal-fold vibration. VOT is negative when vibration of the vocal cords precedes the release of the plosive; VOT is positive when this vibration follows the release of the plosive (Gussenhoven and Jacobs, 1998; Kehoe, Lleó and Rakow, 2004).

Taylor (1975), as cited in Jessen (1998) found longer VOT values for /*p, t, k*/ than /*b, d, g*/ in utterance-initial position in German.[19] Braunschweiler (1997) reported VOT values of voiceless stops over three times longer than VOT values in voiced stops in word-medial position.[20] VOT values in both voiced and voiceless stops had positive values, i.e., the vibration followed the release of the plosive. (see Table 7)

---

[19] Note that the so-called voiced stops are only voiced in intervocalic position, and devoiced at word-initial position. In this position /*p,t,k*/ and /*b,d,g,*/ are distinguished by their VOT values (Braunschweiler, 1997).

[20] Braunschweiler (1997) reported two values per type of stop (voiced vs. voiceless), one value relative to the release before long vowels, and another before short vowels. I calculated the mean, as in Kehoe, Lleó and Rakow (2004).

Table 7. VOT duration means in German (from Braunschweiler, 1997)

	Vowel length	Voiceless stops	Voiced stops
VOT (Release)	Mean after /a/	46 msec	15 msec
	Mean after /a:/	55 msec	16 msec
	Mean	**51 msec**	**16 msec**

Castañeda (1986) studied VOT durations of word-initial stops in Spanish and found near simultaneous release of the plosive and vibration in voiceless stops, resulting in positive (though close to zero) VOT values. In voiced stops, vibration preceded release, resulting in negative VOT values.[21] (see Table 8) Realization of stops in Basque is similar to Spanish, with some minor dialectal variations (Hualde, 2003).

Table 8. VOT duration means in Spanish (from Castañeda, 1986)

	Voiceless stops			**Voiced stops**		
	/p/	/t/	/k/	/b/	/d/	/g/
Mean VOT	6.5 msec	10.4 msec	25.7 msec	-69.8 msec	-77.7 msec	-58 msec
Mean VOT	**14 msec**			**-69 msec**		

Hence, VOT values of voiceless consonants in Spanish tally with VOT values of voiced stops in German. Consequently, it could be the case that German voiced consonants were perceived by Spanish and Basque listeners as similar to their voiceless consonants. Crucially, VOT values of voiceless stops in German are much superior to VOT values of these stops in Spanish. As a result, German voiceless stops could sound to Spanish and Basque native listeners as different from their native voiceless stops. These stops must sound extremely hyperarticulated, almost as voiceless consonants pronounced in emphatic speech or contrastive focus. The onset of emphatic speech or contrastive focus creates a boundary or break between the preceding material and the portion being emphasized. Hence, it could be hypothesized that native listeners of

[21] Castañeda (1986) reported the mean VOT value per each stop, and not a mean value for voices vs. voiceless stops. As in Kehoe, Lleó and Rakow (2004), I calculated the means per type of stop.

Spanish or Basque interpret the German consonants as the initial boundary of a linguistic unit of some sort, or at least interpret them as a set off from the preceding material.

Differences in VOT values of stops in German and French might also explain the frequent-final segmentation preference obtained in the group of French monolinguals. VOT values of stops in French are very similar to the VOT values reported in Spanish. Caramazza and Yeni-Komshian (1974) reported that voiced consonants in French are typically produced with negative VOT values, whereas voiceless stops are typically produced with positive small VOT values.

Only the infrequent syllables in present experiments' artificial language contained the stops /*p ,t, k*/ (see Table 9). If, as proposed, the long VOT values characteristic of German voiceless stops led Spanish, Basque and French native listeners to perceive them as emphatic initial syllables, participants would consequently have segmented the ambiguous stream at the onset of these syllables, inducing a frequent-final segmentation bias.

Table 9. Experiment A. Lexicon of the CV syllables

Frequent syllables	Infrequent syllables
fi, nu, ge	ru, **pe**, du, ba, fo, de, **pa**, ra, **to**, fe, **ta**, **pi**, be, bu, **ko**, mo, **po**, **pu**, mu, ri, **ku**, bo, bi, do, **ka**, na, ro

Last, the frequent-final segmentation preference found in the group of English monolinguals might also result from differences in VOT values between German and English stops. However, the voiced/voiceless distinction of stops is somewhat more complex in English than in French, Basque and Spanish. Voiced stops /*b,d,g*/ are only realized as voiced in medial positions [*b, d, g*], whereas these stops are typically devoiced at word-initial position [*b̥, d̥, g̊*] (Ladefoged, 2001; Lisker and Abramson, 1964). Additionally, the voiceless stops /*p, t, k*/ have two allophonic realizations in English: (a) a aspirated realization [p^h, t^h, k^h] at the onset of stressed syllables, and (b) an unaspirated allophone [*p, t, k*], at other locations (Kahn, 1976, Lisker and Abramson, 1964). Indeed, it has been argued that the distinction between the devoiced initial stops [*b̥, d̥, g̥*] and the unaspirated medial [*p ,t, k*] is rather unclear (Keating, 1984).

In terms of VOT values, English stops, as their German counterparts, typically have positive VOT values. However, VOT values of voiceless stops /*p, t, k*/ differ in their aspirated versus unaspirated allophones (Jansen, 2007). Thus, unaspirated stops [*p, t, k*], have VOT values smaller than 35 msec, resembling VOT values of voiced stops in German. Aspirated stops [p^h, t^h, k^h], on the other hand, are realized with longer VOT values, tallying with VOT values of German voiceless stops. Lisker and Abramson (1969) report a mean VOT of aspirated stops of 69.33 msec.[22] As mentioned, occurrence of the aspirated allophone of /*p, t, k*/ is restricted to the onset of stressed syllables (Kahn, 1976; Lisker and Abramson, 1964). As presented in Chapter 2, section 2.2.1, English is a stress-timed language characterized by the opposition of strong (i.e., stressed) and weak (i.e., unstressed) syllables. Most English words start with a strong syllable (Cutler and Carter, 1987). Crucially, English listeners have been shown to segment speech at the onset of strong syllables (Butterfield and Cutler, 1988; Cutler and Butterfield, 1992; Cutler and Norris, 1988). It is thus plausible that, due to their long VOT values, English participants interpreted the German voiceless stops as English aspirated voiceless consonants, hence taking them as starting a stressed syllable, which generally are word-initial. These consonants might thus have led to frequent-final parsing of the ambiguous grammar. Allophonic variation has been proven a useful cue to speech segmentation in adults (Nakatani and Dukes, 1977; Smith and Hawkins, 2000).

[22] As in Castañeda (1986), Lisker and Abramson (1969) report a mean value per each stop. I calculated the mean for voiced stops as a group.

Summary of findings

The results obtained in this experiment can be summarized into the following conclusions:

(i) Bilingual speakers are able to deploy two processing strategies characteristic of the two languages they command, and switch between these strategies as required by the input.

(ii) The absence of a context effect in L1Spanish bilinguals might result fom differences in language proficiency and language use between this group and the group of L1Basque participants. Alternatively, frequent-initial segmentation might be the unmarked segmentation strategy, and L1Spanish bilinguals might thus not have acquired the marked segmentation strategy, i.e., frequent-final.

(iii) The acoustic-phonetic properties of the input seem to play a decisive role in the participants' segmentation preference of the ambiguous input.

(iv) Adult listeners track the frequency distributions of functor-like elements in an artificial language, elements which are used to segment the input into phrase-like units. Hence, the natural languages' salient frequency of functors leads listeners to treat these frequent elements as anchoring points to syntactic structure, confirming the results in Gervain (2007).

(v) Speakers of VO and OV languages build differing frequency-based representations of the relative order of functors (frequent) and content words (infrequent) elements and these representations influence their segmentation of novel input (i.e., an artificial language), as predicted by Gervain (2007).

Chapter 4. The interplay of statistical and segmental cues in adult speech segmentation

4.1 Introduction

Functors are salient elements in speech, due to their high frequency of occurrence and their tendency to appear at the edges of syntactic phrases. These characteristics turn functors into cues to syntactic structure, which help speakers segment the input into phrases and locate syntactic regularities (Green, 1979; Valian and Coulson, 1988). Gervain (2007) proposes that, due to the correlation between the relative order of content words and functors and the order of verbs and objects present in natural languages (Gervain, 2007; Gervain et al., submitted; Greenberg, 1963), computation of the relative frequency of functors allows infants to build a rudimentary representation of the basic word order of their language. The results reported in Gervain (2007) and in the experiment discussed in the previous chapter (which will be henceforth referred to as Experiment A) reveal that this frequency-based mechanism is also available to adult speakers, who deploy it to segment novel linguistic input (e.g., an artificial language) into phrase-like units.

In Experiment A, speakers of various VO and OV languages were presented with an ambiguous artificial language characterized by alternation of frequent and infrequent elements. The ambiguous language allowed two segmentations: (a) a frequent-initial segmentation that mirrored the order of functors and content words characteristic of VO languages and, (b) a frequent-final segmentation that mirrored the order of functors and content words in OV languages. As predicted by the frequency-based bootstrapping hypothesis, speakers of Basque (OV language) chose a frequent-final segmentation of the ambiguous stream more often than speakers of Spanish, French and English (all VO languages).

In spite of the differences in segmentation preferences found between speakers of VO and OV languages, all groups in Experiment A displayed a bias toward frequent-final segmentation. Thus, while speakers of VO languages (i.e., French, Italian) in Gervain (2007) showed a frequent-initial segmentation preference, though not significantly different from chance, the speakers of French, Spanish and English tested in Experiment A displayed a frequent-final segmentation preference, even though it was weaker than that of Basque speakers.

In the discussion section of the previous chapter, I have suggested that the acoustic-phonetic characteristics of the German voice used in the familiarization and stimuli in Experiment A caused this general frequent-final bias, specifically the VOT (Voice Onset Time) values characteristic of German voiceless stops (Braunschweiler, 1997). Since Gervain (2007) synthesized the stimuli into a Spanish voice (es1, MBROLA database) and stimuli in Experiment A were synthesized with a German voice (de6, MBROLA database), the acoustic difference is a plausible factor behind the difference in the results obtained.

Indeed, current research on cues to speech segmentation reveals that, though adult listeners are able to make use of numerous cues to segment speech (e.g., prosodic cues, acoustic-phonetic cues, statistical cues), they do not weigh all available segmentation cues equally. As presented in Chapter 2, section 2.3, studies that pitted various cues against each other show that these cues are hierarchically organized depending on the relative weight that listeners ascribe to them (Fernandes, Ventura and Kolinsky, 2007; Finn and Hudson Kam, 2008; Mattys, White and Melhorn, 2005, Shukla, Nespor and Mehler, 2007). Mattys, White and Melhorn (2005) propose that the deployment of cues can be ranked within a three-tiered structure. Thus, word recognition in adult listeners is primarily driven by lexical cues (tier I). In the absence of such cues (e.g., in artificial languages) adults rely on segmental cues (tier II) to segment the input, such as phonotactic constraints and acoustic-phonetic cues (e.g., coarticulation and allophonic variation). Only when both lexical and segmental cues are unavailable or ambiguous/uninformative, do adults rely on rhythmic cues (tier III) to segment speech.

Though Mattys, White and Melhorn, (2005) do not include them in the hierarchy, statistical cues have been shown to be outranked by segmental cues such as coarticulation and phonotactic constraints in adult speech segmentation (Fernandes, Ventura and Kolinsky, 2007; Finn and Hudson Kam, 2008). Hence, statistical cues seem to occupy a lower tier of the hierarchy than segmental cues. This is revealed by the study in Fernandes, Ventura and Kolinsky (2007), who presented adult speakers of European Portuguese with an artificial language that contained conflicting coarticulatory and statistical cues to boundaries to word-like units and found that participants preferred word-like units signalled by coarticulatory information, rather than word-like units as cued by statistical information (Transitional Probabilities, TPs). Similarly, Finn and Hudson Kam (2008) pitted phoneme TPs against phonotactic cues and observed that the occurrence of phonotactically illegal sequences prevented their adult participants from tracking the phoneme TPs present in an artificial language, i.e., participants did not choose the statistically defined word-like units over nonword-like units when these units did not obey the phonotactic constraints of their native language.

This strongly suggests that the results of Experiment A are due to the fact that acoustic-phonetic properties of the input influence the segmentation of phrase-like units from the artificial language. Furthermore, the general frequent-final bias, in conjunction with the differences in segmentation preferences found between speakers of VO and OV languages suggest the integration of the two cues, namely, acoustic-phonetic and statistical cues. Research has revealed that none of the available cues in isolation suffices in accurate speech segmentation (Butterfield and Cutler, 1988; Christiansen, Allen and Seidenberg, 1998; Johnson, Jusczyk, Cutler and Norris, 2003). Consequently, successful segmentation depends on the listener’s ability to integrate the available cues (Christiansen, Allen and Seidenberg, 1998). McQueen (1998) investigated the role of rhythm and phonotactic constraints in word segmentation with adult speakers of Dutch and found that, though listeners weighed phonotactic constraints more heavily than rhythmic cues (i.e., stress), rhythm nevertheless modulated the participants’ performance. Similarly, Fernandes, Ventura and Kolinsky (2007) conducted an artificial language learning experiment and observed that congruent statistical and coarticulatory information enhanced participants' performance. Cue integration is crucial during acquisition, due to the absent or scarce lexical knowledge. Infants start integrating the earliest available cues (i.e., rhythmic and statistical cues) around 8 to 9 months of age (Morgan, 1994; Morgan and Saffran, 1995).

In this chapter, I discuss a second artificial language learning experiment (henceforth referred to as Experiment B). In this experiment, the ambiguous artificial language had the same structure as the language in Gervain (2007) and Experiment A, and was synthesized with the Spanish-based voice used by Gervain (2007). However, a new lexicon of CV syllables was created and the duration of the familiarization was reduced from 17 minutes and 30 seconds to 9 minutes and 3 seconds.

The first aim of this experiment is twofold:

(a) to examine whether this change in the acoustic-phonetic properties of the language eliminates the general frequent-final bias found in Experiment A, and

(b) to observe whether the different segmentation preferences predicted by Gervain's (2007) frequency-based bootstrapping hypothesis arise between speakers of VO and OV languages.

Hence, the general aim of this experiment is to examine the interplay between statistical and acoustic-phonetic cues in adult segmentation. To that end, the present experiment tests the segmentation preferences of Basque (L1Basque/L2Spanish bilinguals) and Spanish (Spanish monolinguals and L1Spanish/L2Basque bilinguals) native speakers.

The second aim of this experiment is to further examine the robustness of the context language effect found in bilinguals' performance in Experiment A.

Recall that one of the questions that Experiment A wanted to answer within the framework of the frequency-based bootstrapping hypothesis, was whether the language in which highly proficient bilingual speakers of Basque (OV) and Spanish (VO) received the instructions during the experiment modulated their segmentation preferences of the ambiguous artificial language. To that end, four groups of Basque (OV) / Spanish (VO) bilinguals were tested, sorted by their first language (Basque or Spanish) and the context language, i.e., the language of the instructions (Basque or Spanish). Results showed that bilingual speakers can deploy the segmentation strategies characteristic of their two languages, i.e., their segmentation preference changed as required by depending on the language that was used to issue instructions regarding the experiment. Thus, the group of L1Basque bilinguals that received instructions in Basque chose the frequent-final segmentation significantly more often than the group of L1Basque bilinguals addressed in Spanish during the experiment.

However, no such difference was found in the group of L1Spanish/L2Basque bilinguals. I argued that this asymmetry plausibly originated in differences in proficiency, as a result of the different language use habits that characterized the groups of L1Basque and L1Spanish participants, as revealed by the linguistic background questionnaires collected from the participants. I proposed that these divergent language use habits are a consequence of the particular situation of diglossia found in the Basque Country and argued that the prevailing status of Spanish in the Basque Country entailed two consequences regarding the bilingual population:

(a) A greater habit of language change in L1Basque bilinguals, in order to interact with the vast Spanish monolingual population in the Basque Country (50.8% according to the latest sociolinguistic survey[23]).

(b) A more limited use of Basque by L1Spanish bilinguals, usually restricted to the educational and work areas, resulting in a lower proficiency in Basque as compared to the L1Basque populations' proficiency in their L2, Spanish.

I argued that differences in proficiency between L1Basque and L1Spanish bilinguals could yield different processing of the artificial language. Research on lexical access with highly proficient and lower proficient bilinguals has revealed that highly proficient speakers show symmetrical language switching costs between their L1 and a highly proficient L2 and even between their L2 and a less proficient L3 and between their L1 and L3, whereas lower proficient bilinguals show asymmetrical switching costs between their L1 and L2 (Costa and Santesteban, 2004; Costa, Santesteban and Ivanova, 2006; Meuter and Allport, 1999).

Alternatively, I suggested that the different performance in the L1Basque and L1Spanish bilinguals might be explained if the frequent-initial segmentation characteristic of VO languages was the unmarked segmentation strategy, and the frequent-final segmentation characteristic of OV languages the marked segmentation strategy. L1Basque bilinguals, unlike L1Spanish bilinguals, would have acquired the marked strategy and hence be able to make use of the two segmentation strategies.

[23] V Sociolinguistic Survey 2011, Basque Government.

In Experiment B, four new groups of L1Basque/L2 Spanish bilinguals are tested, in order to examine whether the context language effect found in Experiment A is replicated in the new artificial language.

4.2 Methodology

4.2.1 Participants

119 graduate students (92 females) from the University of the Basque Country (UPV/EHU), with a mean age of 21.22 (age range 18 to 41) took part in this experiment. Participants were sorted into five groups, depending on their native language and the language in which they received the instructions (i.e., context language). 27 Spanish monolinguals (21 females) took part in the experiment, with a mean age of 25.19 years (age range 18 to 41). 27 L1Basque/L2Spanish bilinguals (19 females, mean age 20.44, age range 18 to 31) participated in the experiment receiving the instructions in Basque and another 23 L1Basque/L2Spanish bilinguals (17 females, mean age 18, age range 18 to 18) were addressed in Spanish. Similarly, 21 L1Spanish/L2Basque bilinguals (15 females, mean age 20.62, age range 18 to 26) took part in the experiment, conducted in Basque and another 21 L1Spanish/L2Basque bilinguals (20 females, mean age 21.24, age range 18 to 27) received the instructions in Spanish.

The linguistic background of the participants was collected by means of the same questionnaire described in the previous experiment. Both groups of bilinguals acquired their L2 at around 3 to 5 years of age. All L1Spanish/L2 Basque participants (a) had completed all their studies in Basque and/or were conducting studies in Basque at the UPV/EHU, (b) possessed a degree in Basque studies, or (c) had the EGA language certificate (Basque Language Certificate, C1 level in the Common European Framework). Participants were naïve to the purpose of the experiment and received 4€ in compensation for their participation.

4.2.2 Materials

As mentioned above, the structurally ambiguous artificial language in the present experiment shared the same structure with the artificial language in Gervain (2007) and Experiment A. Thus, the basic unit of this language was again six-syllable long, with an *aXbYcZ* structure. *a*, *b* and *c* represented frequent categories and contained one CV syllable token each. *X*, *Y* and *Z* represented infrequent categories and contained 9 CV syllable tokens each. Again, the new language comprised a total of 3 frequent tokens and 27 infrequent tokens, and frequent categories were nine times more frequent than infrequent categories. (see Table 10)

Table 10. Structure of the hexasyllabic unit: aXbYcZ

	Frequent categories	**Infrequent categories**
Structure	a, b, c	X, Y, Z
Number of tokens	3	27
Number of tokens per category	a = 1; b = 1; c = 1	X = 9; Y = 9; Z = 9
Syllabic structure of the categories and tokens	CV	CV
Relative frequency of the categories	Frequent categories: 9-times more frequent than infrequent categories.	

A new lexicon of CV syllables was created for the present language, as shown in Table 11.

Table 11. Experiment B. Lexicon of the CV syllables

a	X	b	Y	c	Z
fi	fe	nu	fu	pe	fo
	ka		ko		ki
	lu		li		ku
	mu		lo		le
	ne		mo		ma
	pa		na		pi
	so		pu		po
	ti		sa		ta
	to		te		tu

Phonotactic constraints were taken into consideration in order to create the lexicon. Thus, all CV syllables that contained vibrants were excluded from the inventory of CV syllables in Experiment B. In Basque, unlike in Spanish, vibrants are phonotactically illegal in word-initial position, except for some recent borrowings from Spanish (Hualde, 2003). Hence, the syllables *ra, ri, ro* and *ru*, present in the lexicon in Experiment A, have extremely low syllabic frequencies at word-initial position,[24] as shown by the online application SYLLABARIUM[25] (Duñabeitia, Cholin, Corral, Perea and Carreiras, 2010). As presented in Chapter 2, section 2.3, adult speakers are sensitive to phonotactic constraints and make use of these constraints in speech segmentation (McQueen, 1998; van der Lugt, 2001; Weber, 2001). It is plausible that L1Basque bilinguals might treat syllables that contain a vibrant as non-initial, which would create a bias in the segmentation preference of this group. Though L1Basque participants in Experiment B were all highly proficient speakers of Spanish and bilinguals appear to make use of the phonotactic constraints of their L2 , interference of the L1 phonotactic constraints in the processing of input in the L2 has been found in a number of studies (Alternberg and Cairns, 1983; Flege and Wang, 1989; Sebastián-Gallés and Bosch, 2002; Trapman and Kager, 2009; Weber and Cutler, 2006).

[24] Type frequencies: *ra* = 6; *ri* = 1; *ro* = 5; *ru* = 1. Token frequencies: *ra* = 12; *ri* = 1; *ro* = 5; *ru* = 1.
[25] http://www.bcbl.eu/syllabarium/index.php

Also, voiced stops were excluded from the lexicon of CV syllables. Voiced stops have two allophonic realizations both in Spanish and Basque: (a) as a stop after a pause and after a nasal consonant (and after /*l*/ for the voiced stop /*d*/), and (b) as an approximant at other positions (Hualde, 2003; Navarro Tomás, 1968). No MBROLA voice available in Spanish includes this distinction. Allophonic variation is one of the available cues that assist adults in word segmentation, as shown in Chapter 2, section 2.3 (Lehiste, 1960; Nakatani and Dukes, 1977; Smith and Hawkins, 2000). Nine of the syllables that formed the lexicon in Experiment A contained the voiced stops /*b, d, g*/ (i.e., *du, ba, de, be, bu, ge, bo, bi, do*). Hence, these syllables were excluded from the lexicon in Experiment B in order to ensure the maximum naturalness of the stimuli and avoid potential allophonic effects.

The six-syllable unit was concatenated 377 times to create the familiarization string. The order of the frequent categories remained constant (with the structure ***fi****X****nu****Y****pe****Z*). Meanwhile, the occurrence of the tokens of each infrequent category (*X*, *Y* and *Z*) was randomized. Occurrences of the sequences that would constitute the test items were excluded from the resulting string.

Familiarization was 9 minute and 3 second long, hence shorter than the 17 minute 30 and second long familiarization in Gervain (2007) and Experiment A. In order to ensure ambiguity, the first and last minute of the familiarization were ramped in intensity using the program Audacity.[26] The stream allowed for two types of segmentation: (a) a frequent-initial segmentation (*aXbYcZaXbYcZ*…), or (b) a frequent-final segmentation (*XbYcZaXbYcZa*…). (see Table 12)

Table 12. Experiment B. Familiarization

Segmentation preference	Structure	Familiarization
	…**a**X**b**Y**c**Z**a**X**b**Y**c**Z**a**X**b**Y**c**Z **a**X**b**Y**c**Z**a**X**b**Y**c**Z**a**X**b**Y**c**Z…	…**fi**ne**nu**li**pe**ki**fi**ka**nu**lo**pe**le**fi**ka**nu**li **pe**pi**fi**pa**nu**fu**pe**ku**fi**ne**nu**na**pe**pi**fi**…
Frequent-final	…**a[***XbYcZa***]**X**b**Y**c**Z**a**X**b**Y…	…**fi[***nenulipekifi***]**ka**nu**lo**pe**le**fi**ka**…**
Frequent-initial	…**a**X[***bYcZaX***]**b**Y**c**Z**a**X**b**Y…	…**fi**ne**[***nulipekifika***]nu**lo**pe**le**fi**ka**…**

[26] http://audacity.sourceforge.net/

As in Experiment A, test stimuli consisted of 36 six-syllable long items. 18 of the test items instantiated a frequent-initial order, i.e., started with a frequent category (e.g., *aXbYcZ*: *fiKAnuFUpePO*). The remaining 18 test items displayed a frequent-final order, i.e., started with an infrequent category (e.g., *XbYcZa*: *TInuNApeKUfi*). Within test items all categories (frequent and infrequent) occurred with similar frequency in all possible positions. Thus, 6 of the 18 frequent-initial items started with the syllable *fi*, i.e., token of the category *a* (*aXbYcZ*), another 6 test items started with the syllable *nu*, i.e., token of the category *b* (*bYcZaX*) and the remaining 6 test items started with the syllable *pe*, i.e., token of the category *c* (*cZaXbY*). A similar distribution characterized the frequent-final elements.[27] Each infrequent syllable occurred 4 times in the test trials, i.e., twice in the frequent-final items and twice in the frequent-initial items. Additionally, all sequences of two or more syllables that were potential words in Spanish and Basque were avoided.

The items and familiarization were synthesized with the es1 (Spanish male) voice of the MBROLA database (Dutoit, 1997), i.e., the same voice used in Gervain (2007), instead of the German voice (de6, German male) used in Experiment A. As in Experiment A, all phonemes in both familiarization and test items were 120 msec long and had a monotonous f0 of 100 Hz. To ensure f0 uniformity in the materials, any deviation from the intended 100 Hz was corrected using the software PRAAT (Boersma and Weenink, 2012).

As in Experiment A, during the experimental task test items were presented in 36 pairs, each pair containing two items with opposite orders: a frequent-final item and a frequent-initial item. The order of the items was counterbalanced. Each of the 36 experimental test items appeared twice, one as the first member of a pair, and another as the second member of another pair. Test items in each test pair were separated by a 700 msec pause. The items were counterbalanced. Thus, from the 6 test items that started with the frequent syllable *nu-*, two were paired with test items that ended with *-nu*, another two were paired with test items ending with the frequent syllable *-pe* and the remaining two were paired with test items ending in the frequent syllable *-fi*. Pairing of the remaining test items followed the same distribution, as shown in Appendix 2. The order of presentation of the test pairs was pseudorandomized, i.e., the same item never appeared in consecutive trials.

[27] See Appendix 2 for the complete list of test stimuli.

4.2.3 Procedure

All experiments were conducted in the psycholinguistics laboratory at the UPV/EHU. Participants were tested individually in a quiet room. The experiment was displayed in a computer screen and participants were provided with a set of high quality headphones. The procedure was hence identical to the one described in Experiment A.

4.3 Results

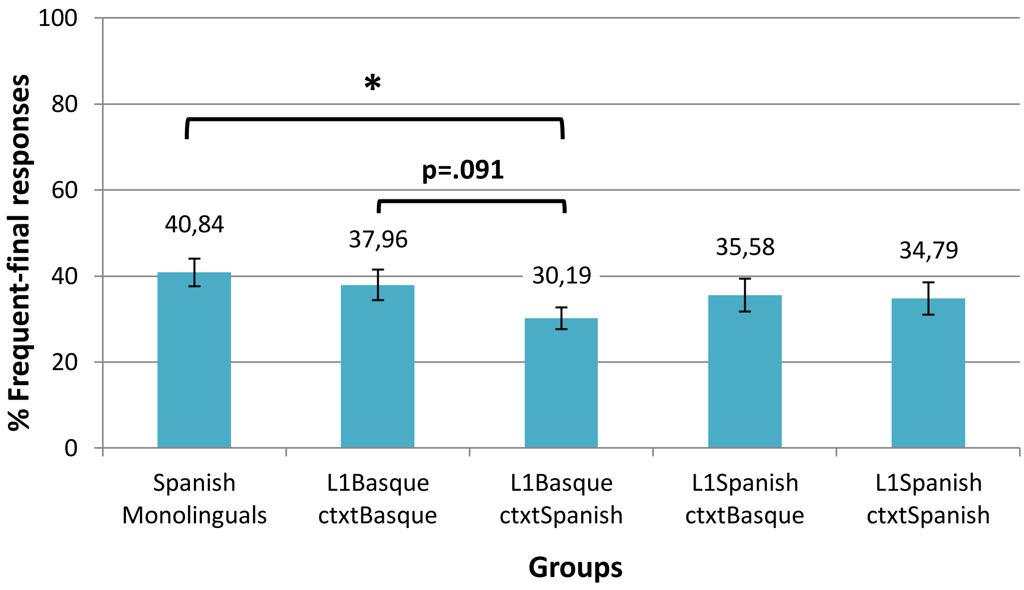
Data analysis was conducted using the software SPSS. All groups showed a general preference for frequent-initial segmentation of the ambiguous language, as shown in Table 13 and Graph 7.

Table 13. Experiment B. Mean of frequent-final preference per group

Group	Frequent-final responses (out of 36)	Percentage of frequent-final responses
Spanish monolinguals	14.7	40.84%
L1Basque context Basque	13.67	37.96%
L1Basque context Spanish	10.87	30.19%
L1Spanish context Basque	12.81	35.58%
L1Spanish context Spanish	12.52	34.79%

Kolmogórov-Smirnov tests revealed normal distribution of the dependent variable in all groups. One-sample t-tests were conducted with each group. The mean score of the sample was compared to a test value of 18 (out of 36, a value of 18 would amount to chance). Segmentation preferences in all groups differed significantly from chance (all $p \leq .009$).

Graph 7. Experiment B. Percentage of frequent-final responses per group

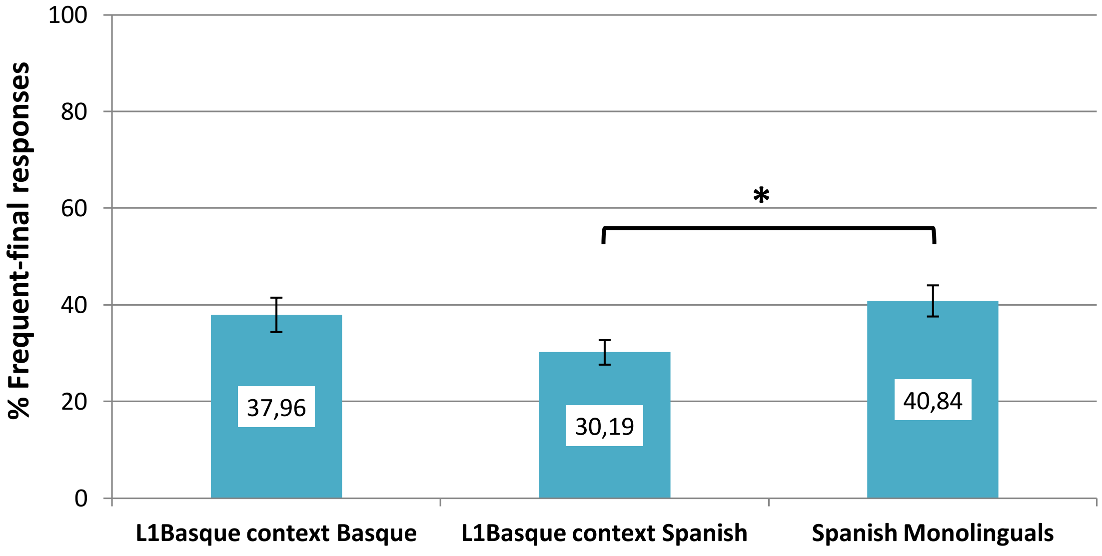


A one-way ANOVA with the factor *group* was conducted, which revealed no significant main effect (F(4.114) = 1.401, p = .238). Since the present experiment aimed to specifically examine (a) the segmentation preferences of the bilinguals participants, as well as (b) the segmentation preferences of speakers of VO and OV languages, pair-wise comparison of the groups was conducted, notwithstanding the absence of a main effect revealed by the ANOVA.Pair-wise comparisons of the groups were conducted by means of separate independent sample t-tests.

4.3.1 OV vs. VO languages

Pair-wise comparisons revealed no significant differences between the group of L1Basque context Basque (M = 13.67, SD = 6.66) and the group of Spanish monolinguals (M = 14.7, SD = 5.02) (t(52) = .600, p = .551). However, the group of Spanish monolinguals chose a frequent-final segmentation significantly more often than the group of L1Basque context Spanish (M = 10.87, SD = 4.36) (t(48) = 2.54, p = .014).

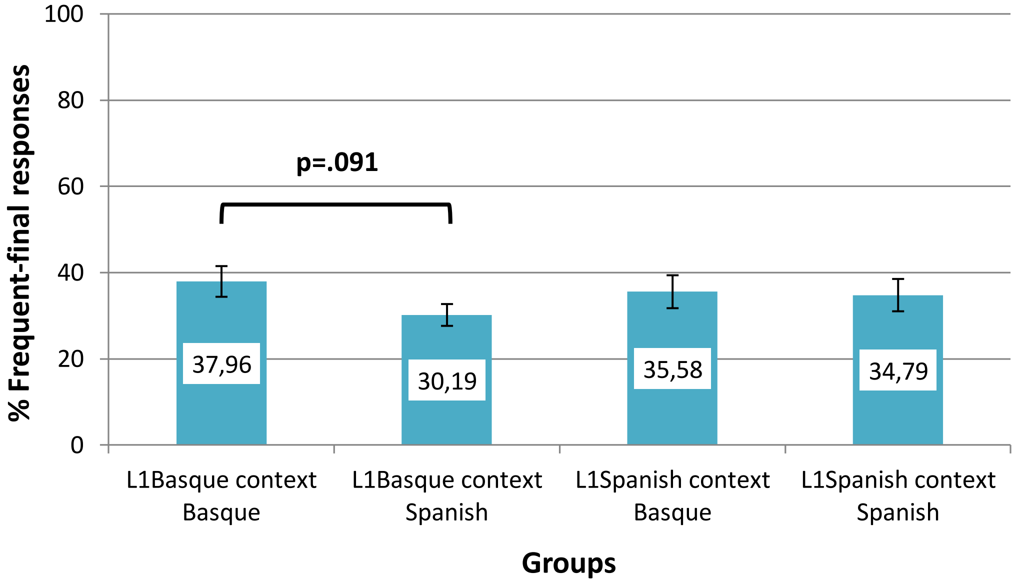
Graph 8. Frequent-final preference of L1Basque bilinguals vs. Spanish monolinguals



4.3.2 Context language

In order to examine the potential influence of the context language in the bilinguals' segmentation preferences, pair-wise comparisons were conducted within the bilingual participants. The group of L1Basque context Basque (M = 13.67, SD = 6.66) chose a frequent-final segmentation more often than the group of L1Basque context Spanish (M = 10.87, SD = 4.36). However, this difference was only marginally significant (t(48) = 1.723, p = .091). The groups of L1Spanish context Basque (M = 12.81, SD = 6.31) and L1Spanish context Spanish (M = 12.52, SD = 6.19) showed a non-different segmentation preference (t(40) = .148, p = .883).

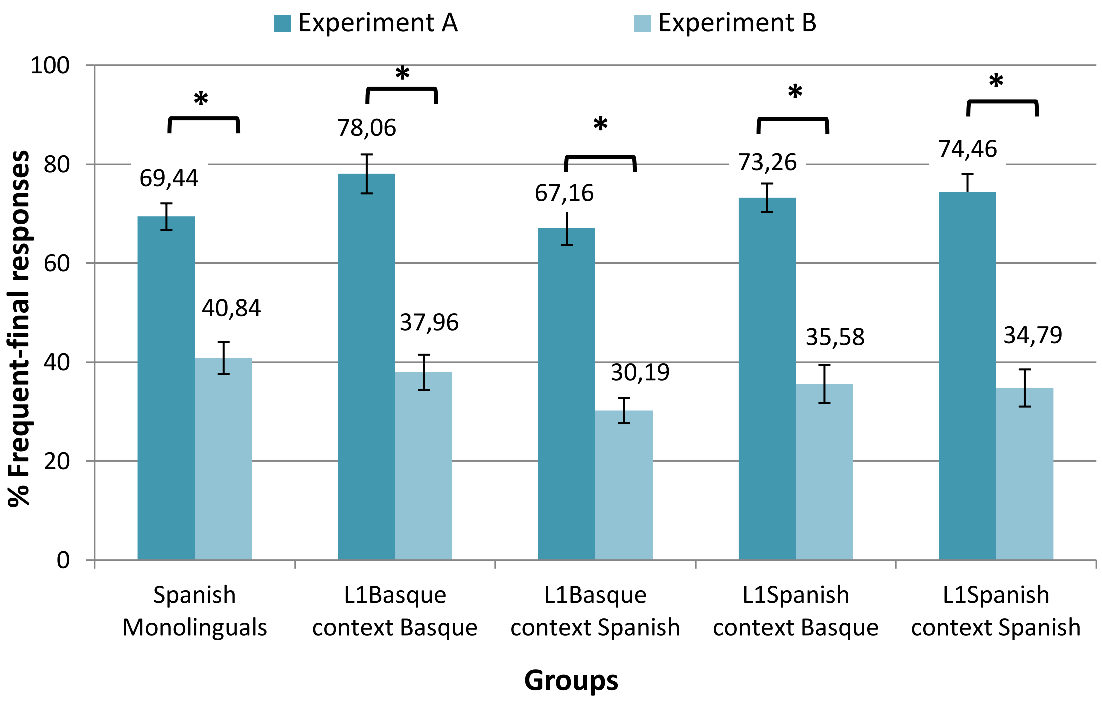
Graph 9. Context languages' influence in segmentation preferences



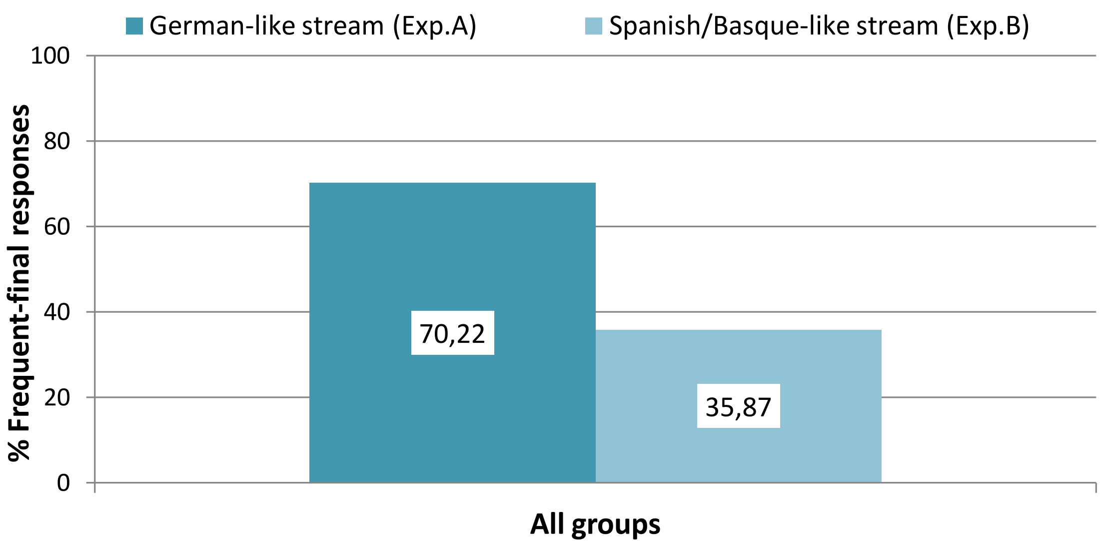
4.3.3 Experiment A vs. Experiment B

Pair-wise comparisons were conducted for each group between the segmentation preference in this experiment and the preference obtained in Experiment A. The analyses revealed significant differences across experiments in all five groups (all $p < .001$). Thus, participants in Experiment A displayed a frequent-final segmentation preference, whereas participants in Experiment B displayed a frequent-initial segmentation preference. This result suggests that segmental cues were decisive in the segmentation of the stream by the adult participants.

Graph 10. Comparison of frequent-final responses in experiments A and B



Graph 11. Comparison of general means in experiments A and B



As described in the methodology section, the familiarization stream in Experiment B was shorter (9 minutes and 3 seconds) than the familiarization stream in Experiment A (17 minutes and 30 seconds). In order to rule out this change in duration as the cause of the reversed segmentation preference found across the two experiments, a control experiment was conducted.

4.4 Control Experiment

4.4.1 Methodology

This experiment was an exact replica of Experiment B, except for one manipulation, i.e., a new familiarization was designed with a duration of 17 minutes and 3 seconds (as in Experiment A). If reduction of the manipulation yielded a reversal in segmentation preference, similar results to Experiment A are expected in the present control. If, on the other hand, the duration of the familiarization phase did not influence the segmentation preference of the participants, no difference is expected between the results of Experiment B and the present control.

4.4.1.1 Participants

13 undergraduate students (10 females) from the University of the Basque Country (UPV/EHU), with a mean age of 18.92 (age range 18 to 22) took part in this experiment. Participants formed a sole group of L1Basque/L2Spanish bilinguals. Participants were addressed and received the instructions in Basque. These participants acquired Spanish around 4 years of age. The linguistic background of the participants was collected by means of the questionnaire reported in the previous experiments. Participants were naïve to the purpose of the Experiment and received a 4€ compensation for their participation.

4.4.1.2 Materials and procedure

The design of the materials and procedure were replicas of Experiment B, with the sole exception of the increase of the duration of the familiarization phase from 9 minutes 3 and seconds to 17 minutes and 4 seconds.

4.4.2 Results

Data analysis was conducted using the software SPSS. The Kolmogórov-Smirnov test revealed a normal distribution of the dependent variable. A one-sample t-test was conducted, in which the mean score of the sample was compared to a test value of 18 (out of 36, a value of 18 would amount to chance). The control group of L1Basque context Basque ($M = 16.23$, $SD = 9.65$) chose a frequent-initial segmentation more often than a frequent-final segmentation. However, this preference did not differ significantly from chance ($t(12) = -.661$, $p = .521$).

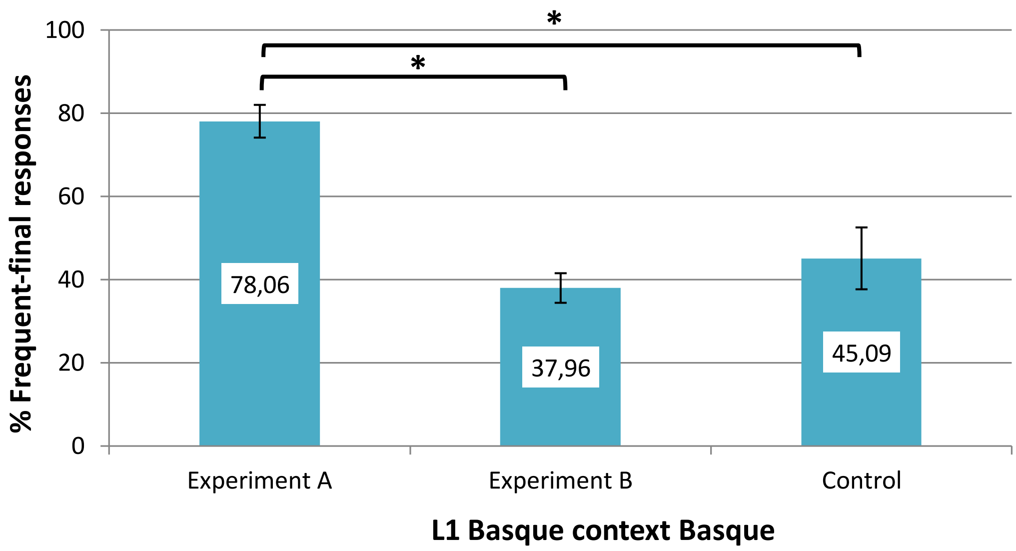
Table 14. Control experiment. Frequent-final preference per group

Group	Frequent-final responses (out of 36)	Percentage of frequent-final responses
L1Basque context Basque	16.23	45.09%

4.4.2.1 Experiment A vs. B vs. Control

An ANOVA with the factor *experiment* was conducted with the results of the L1Basque context Basque bilinguals' group across experiments A, B and control, and a significant main effect was obtained ($F(2.57) = 23.707$, $p < .001$). Pair-wise comparisons were conducted by means of independent t-tests. The t-tests revealed a significant difference in the segmentation preference of the L1Basque context Basque group between experiments A ($M = 28.1$, $SD = 6.34$) and control ($M = 16.23$, $SD = 9.65$) ($t(31) = 4.276$, $p < .001$). The groups' segmentation preference in experiments B ($M = 13.67$, $SD = 6.66$) and control did not differ significantly ($t(38) = -.983$, $p = .332$).

Graph 12. Comparison of frequent-final responses in experiments A, B and control



4.5 Discussion

4.5.1 The interplay between segmental and statistical cues

Experiment B aimed to examine whether segmental information superseded statistical information in adult speech segmentation, i.e., whether the general frequent-final segmentation preference obtained in Experiment A resulted from a bias created by the acoustic-phonetic properties of the German voice used to synthesize the stimuli, namely the longer VOT of voiceless stops in German than in Spanish stops. To that end, the artificial language and test stimuli in Experiment B were synthesized with the same Spanish voice (es1, MBROLA database) as in Gervain (2007). A general bias towards frequent-initial segmentation of the ambiguous language was obtained in the present experiment. This shift in segmentation preference, from a general frequent-final segmentation in Experiment A to the general frequent-initial segmentation preference found in Experiment B suggests that the acoustic-phonetic properties of the input did indeed cause the bias observed in Experiment A. Consequently, this result suggests that acoustic-phonetic cues outrank the statistical cues contained in the input in adult speech segmentation, since statistical cues were held constant across experiments A and B.

The design of the artificial language in Experiment B contained three changes with respect to the artificial language used in Experiment A:

(a) The mentioned change in the voice used to synthesize the familiarization and the test stimuli, from a German voice to the Spanish voice used by Gervain (2007).

(b) A change on some of the CV syllables that formed the lexicon of the language.

(c) A reduction of the duration of the familiarization from 17 minutes and 20 seconds to 9 minutes and 3 seconds.

Thus, in addition to the change in the voice used to synthesize the stimuli, the artificial language contained two further changes that have to be ruled out as potential causes for the general frequent-initial bias found in all groups.

The control experiment reported in section 4.4 of the present chapter ruled out the reduction in the duration of the familiarization as the cause of the general frequent-final bias obtained in Experiment B. Recall that the artificial language in this control experiment was an exact replica of the language in Experiment B with one exception, i.e., an increase in the duration of the familiarization to 17 minutes and 3 seconds, hence similar to the duration of the stream in Experiment A. Notwithstanding this increased duration, the group of L1Basque context Basque participants in the control experiment showed a frequent-initial segmentation preference (45.09%[28]) that did not differ from the preference found in the same group in Experiment B (37.96%). Indeed, research on artificial language learning has shown that adult speakers successfully track statistical regularities after a familiarization of around 10 minutes (Onnis, Monaghan, Richmond and Chater, 2005; Toro, Shukla, Nespor and Endress, 2008). Furthermore, regularity extraction in adults has been found even after a 3 minute long exposure to the artificial language (Mersad and Nazzi, 2011).

Similarly, an analysis that compared the frequencies in Spanish and Basque of the CV syllables from experiments A and B ruled out a change in the frequency of the new lexicon of CV syllables as the cause of the general frequent-initial bias.

13 of the 30 CV syllables that formed the lexicon in Experiment A were changed in the design of Experiment B in order to avoid phonotactic and allophonic effects (as described in the methodology). (see Table 15)

[28] The mean values provided henceforth represent percentages of frequent-final responses.

Table 15. Original syllables from Experiment A and new syllables in Experiment B

Syllables in Experiment A	New syllables in Experiment B
ru, du, ba, de, ra, be, bu, ge, ri, bo, bi, do, ro	fu, ki, le, li, lo, lu, ma, ne, sa, so, te, ti, tu

As mentioned above, the syllabic frequencies of the CV syllables in experiments A and B were compared in Spanish and Basque. Syllabic frequencies in Basque and Spanish were taken from the online application SYLLABARIUM (Duñabeitia, Cholin, Corral, Perea and Carreiras, 2010). Analyses were conducted both of token and type frequencies. All frequencies were converted into $\log_{10}$ values for analysis.

The syllabic frequency of the syllables at word-initial position in experiments A and B were compared in each language by means of independent sample t-tests. Analyses of the token and type frequencies of all syllables of the lexicon in Experiment A vs. Experiment B in Basque revealed no significant differences.[29] No significant differences were found either in the analyses of the syllables in Spanish. The syllabic frequencies of the subset of thirteen syllables changed in experiments A and B were similarly analyzed. Analyses revealed no significant differences in token or type frequencies neither in Basque nor in Spanish. SYLLABARIUM does not provide the frequencies of the syllables in word-final position. However, it provides the general token and type frequencies of the syllables in all positions. These frequencies were analyzed in exactly the same way as the syllabic frequencies in word-initial position. No significant differences were found in any of the comparisons conducted.

Hence, these analyses showed that the new repertoire of CV syllables in Experiment B did not entail changes in the syllabic frequencies, as compared to the syllabic repertoire in Experiment A.

The change in the voice used to synthesize the familiarization and test stimuli from a German voice in Experiment A to the Spanish voice used in Gervain (2007) would thus remain as the only difference between both experiments. I claim that this difference appears to have caused the shift in segmentation preference found in experiments A (general frequent-final segmentation) and B (general frequent-initial

[29] For the sake of clarity no statistical results are included in this section. See Appendix 3 for a detailed account of the statistical analysis.

segmentation). In the discussion of Experiment A, I argued that the VOT values that characterize voiced and voiceless stops in German might have biased participants to a frequent-final segmentation preference of the stream. The stream in Experiment B only contained phonemes that belong to the participants' native languages, hence eliminating this acoustic-phonetically induced bias. The shift in segmentation preference found in experiments A and B suggests that segmentation was driven by acoustic-phonetic cues, rather than by the statistical information present in the stream, which remained constant across both experiments. This result provides evidence on the sensitivity to the acoustic-phonetic properties of the input in adults, properties that are available cues that assist listeners in speech segmentation and lexical access (Andruski, Blumstein and Burton, 1994; Fernandes, Ventura and Kolinsky, 2007; Mattys, White and Melhorn, 2005; Shatzman and McQueen, 2006). Furthermore, this result supports a hierarchical approach to the segmentation cues, as proposed by Mattys, White and Melhorn (2005), according to which adult speakers rely on segmental cues such as acoustic-phonetic information, whenever lexical information is unavailable, rather than on statistical or prosodic cues, i.e., acoustic-phonetic cues outrank statistical cues.

4.5.2 Segmentation preferences on OV versus VO languages

According to the predictions of the frequency-based bootstrapping hypothesis (Gervain, 2007) a greater frequent-final segmentation preference was expected in the group of L1Basque participants (OV) which received the context in Basque as compared to native speakers of Spanish (VO) (i.e., Spanish monolinguals and L1Spanish bilinguals). However, no such difference was obtained. Instead, a similar general frequent-initial segmentation preference was found in these groups. This frequent-initial segmentation preference in the speakers of VO languages, i.e., Spanish monolinguals (14.7 out of 36, i.e., 40.84%) and L1Spanish bilinguals context Spanish (12.52 out of 36, i.e., 34.79%), tallies with the frequent-initial preference found in Gervain (2007) with speakers of other VO languages (i.e., French: 16.95 out of 36 and Italian: 17.05).

As mentioned above, the group of L1Basque bilinguals context Basque in Experiment B also showed a frequent-initial segmentation preference (13.67 out of 36, i.e., 37.96%), as opposed to the pronounced frequent-final segmentation preference

(25.75 out of 36) found in a similar group in Gervain (2007), and replicated by the group of L1Basque context Basque in Experiment A (28.1 out of 36, i.e., 78.06%). This frequent-initial preference obtained among L1Basque participants in Experiment B (as opposed to the frequent-final preference obtained in Gervain, 2007) is rather striking, given that Gervain's (2007) language and the artificial language in Experiment B shared the same structure, procedure and, crucially, the same voice. Furthermore, contrary to the predictions of the frequency-based bootstrapping hypothesis, the frequent-final segmentation preference displayed by the group of L1Basque bilinguals context Basque (37,96%) was smaller (though not significantly) than the frequent-final segmentation preference obtained in the group of Spanish monolinguals (40.84%), i.e., the group of L1Basque bilinguals showed a greater frequent-initial segmentation preference than the group of Spanish monolinguals.

A second group of L1Basque participants took part in Experiment A, i.e., L1Basque context Spanish. As the group of L1Basque context Basque, this group of L1Basque context Spanish showed a frequent-initial segmentation preference (30.19%)[30] in Experiment B, as opposed to the frequent-final preference (67.16%) found in a similar group in Experiment A. A significant difference was found between the segmentation preferences of this group of L1Basque context Spanish bilinguals and the group of Spanish monolinguals (40.84%). Again, the group of L1Basque bilinguals displayed a more marked tendency for frequent-initial segmentation than Spanish monolinguals.

A possible cause of the unexpected frequent-initial segmentation preference found in the groups of L1Basque bilinguals might originate from a difference in the linguistic background of the L1Basque participants of experiments A and B. In order to rule out this possibility, I examined the linguistic background questionnaires that the participants filled out during the experimental session. The information contained in the questionnaires was compared to the L1Basque participants' information collected in Experiment A. No differences were found in the linguistic backgrounds between L1Basque participants in experiments A and B. Both groups came mainly from small towns and villages in the province of Gipuzkoa, had been raised in homes were Basque was exclusively spoken and had received their education in Basque. Also, both groups

[30] Note that the mean values provided represent percentages of frequent-final responses.

reported speaking predominantly Basque in situations other than at school/university/work and at home during infancy, adolescence and adulthood.[31]

Another possible (though rather unlikely) explanation for the L1Basque bilinguals' different segmentation preference in Gervain (2007) and in Experiment B might lie on the fact that the lexicon in Experiment B contained two consonants that were not present in the lexicon in Gervain (2007) and Experiment A, i.e., /*s, l*/. These two synthesized consonants might have contained some acoustic information that would have biased the participants towards a frequent-initial segmentation preference. Thus, 7 of the 13 new syllables that formed the lexicon in Experiment B contained consonants that were also included in the lexicon in Gervain (2007) (i.e., /*f, k, m, n, t*/, but the remaining 5 syllables contained the new consonants /*s, l*/. The syllables that contained these two consonants (i.e., *le, li, lo, lu, sa, so*) all formed part of the infrequent CV syllables of the lexicon. Hence, in order to generate a bias toward frequent-initial, the syllables should have contained acoustic properties that signalled a non word-initial position.

A last possible explanation for the frequent-initial segmentation preference obtained in the L1Basque participants might be that this group, as well as the other groups tested (i.e., Spanish monolinguals and L1Spanish bilinguals), might be segmenting the artificial language according to a universally unmarked or default segmentation strategy, i.e., frequent-initial segmentation. I first stated the possibility of a universal segmentation strategy in the discussion of Experiment A, as a plausible explanation for the absence of a context effect observed in the group of L1Spanish bilinguals.

Given that the stream in Experiment B only contained phonemes that belong to the participants' native languages, hence eliminating the acoustic-phonetic information (i.e., the differences in VOT values between German and Spanish/Basque) that biased participants to a frequent-final segmentation preference of the stream in Experiment A, the frequent-initial segmentation might emerge as the universal segmentation preference. However, this hypothesis does not account for the frequent-final segmentation preference obtained in Gervain (2007) with the same Spanish voice used

[31] A more detailed description of the linguistic background of the group of L1Basque bilinguals is presented in the next section. Also, see Appendix 4 for statistical analyses.

in Experiment B, in speakers of Basque and Japanese. The conflicting preference found in Gervain (2007) and Experiment B opens thus the ground for future investigations in order to elucidate the plausibility of a universal frequent-initial segmentation strategy.

4.5.3 Context language

In the present experiment, the group of L1Basque bilinguals which received the instructions in Basque chose a frequent-final segmentation of the ambiguous language more often (37.96%) than the group of L1Basque bilinguals addressed in Spanish (30.19%), though this result was only marginally significant ($p = .091$). No differences were found in the segmentation preferences within the L1Spanish bilinguals (35.58% in context Basque vs. 34.79% in context Spanish). The direction of the difference in their segmentation preferences and the absence of a difference within the L1Spanish participants replicate the picture obtained in Experiment A, notwithstanding the fact that the difference within the L1Basque participants is only marginally significant. Therefore, this context effect found in experiments A and B suggests that the effect is rather robust.

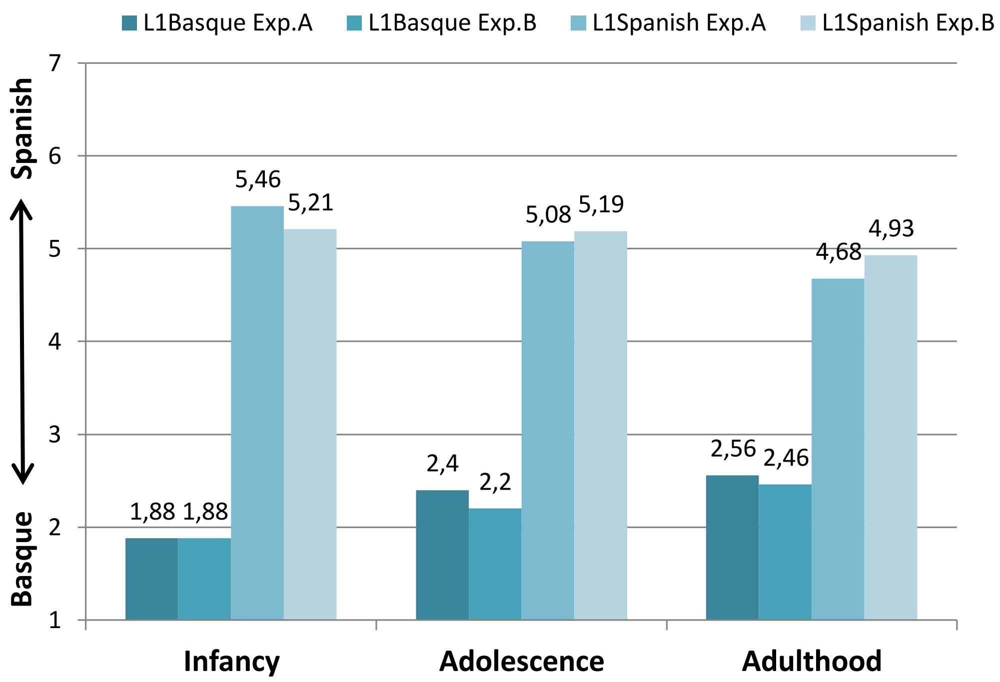
As in Experiment A, analyses of the linguistic background questionnaires were conducted, in order to account for the differences in performance between the L1Basque and L1Spanish bilinguals. Again, only few of the L1Basque participants (5 out of 50) came from province capitals. Also, 31 of the 50 L1Basque participants came from Gipuzkoa, the province with highest concentration of Basque native speakers. The L1Spanish bilinguals, on the other hand, showed the opposite pattern. Thus, half of the 42 participants lived in province capitals. Also, only 7 of the 42 came from Gipuzkoa.

The questionnaire inquired about the frequency of use of their two languages in situations other than at school/university/work and at home, at three points of development: infancy, adolescence and adulthood. Participants rated the frequency of use by choosing one of seven options. The results from the analysis replicated the differences in frequency of use obtained in the analysis of the questionnaires in Experiment A.[32] Thus, no differences on frequency of use of the two languages were found neither within the L1Basque bilinguals (context Basque vs. context Spanish), nor within the L1Spanish bilinguals (context Basque vs. context Spanish). Again,

[32] See Appendix 4 for the results of the statistical analyses.

significant differences were found between L1Basque and L1Spanish bilinguals' frequency of use of their two languages, in all three points of development. Thus, the group of L1Basque bilinguals reported predominant use of Basque in situations other than at school or at home, whereas the group of L1Spanish bilinguals reported using Spanish predominantly in their interactions. Furthermore, comparison of the participants' frequency of use between experiments A and B revealed no significant differences at any of the three stages of development, neither in the responses of the L1Basque nor in the L1Spanish bilinguals.

Graph 13. Frequency of use of Basque and Spanish in experiments A vs. B



The context effect obtained in the L1Basque/L2Spanish bilinguals suggests that bilinguals are able to acquire the segmentation strategies of the two languages they command. This ability to make use of the segmentation strategies of both their L1 and L2 has also been found in the study of other speech segmentation cues, such as statistical, prosodic and segmental (Sanders, Neville and Woldorff, 2002; Sebastián-Gallés and Bosch, 2002; Weiss, Gerfen and Mitchell, 2009). Whereas early and late learners of an L2 seem to achieve native-like or close to native-like levels of processing in some segmentation strategies (e.g, phonotactic constraints), acquisition of other segmentation strategies appears to be more limited (e.g., rhythm) (Cutler, Mehler, Norris and Seguí, 1989; 1992; Sanders, Neville and Woldorff, 2002; Sebastián-Gallés and Bosch, 2002).

Thus, as described in Chapter 2, section 2.1.1, Weiss, Gerfen and Mitchell (2009) presented adult speakers with mixed input of two artificial languages, that contained either congruent or incongruent statistical information (i.e., Transitional Probabilities, TPs, between adjacent syllables), aiming to emulate language acquisition in a bilingual environment and found that the adult speakers separately computed the TPs of adjacent syllables in both artificial languages, regardless of whether the statistical information was congruent or incongruent in the two languages.

Regarding other segmentation cues, several studies have shown that both early and late bilingual speakers are able to learn and make use of the phonotactic constraints of the L2 and that acquisition of L2 phonotactics builds up with proficiency (Altenberg and Cairns, 1983; Sebastián-Gallés and Bosch, 2002; Trapman and Kager, 2009; Weber and Cutler, 2006). However, interference of the L1 phonotactic constraints has been consistently found in the processing of input in the L2 (Alternberg and Carins, 1983; Flege and Wang, 1989; Sebastián-Gallés and Bosch, 2002; Weber and Cutler, 2006). Indeed, the scarce evidence on bilingual infants' sensitivity to phonotactic constraints has revealed an asymmetry in the acquisition of phonotactic constraints in the two languages, i.e., acquisition of the phonotactic constraints of the bilinguals' dominant language precedes acquisition of the phonotactic constraints of the non-dominant language (Sebastián-Gallés and Bosch, 2002).

The scarce research that has examined whether bilingual speakers command the two rhythm-based segmentation procedures characteristic of their two language has revealed that acquisition of the L2's rhythm-based segmentation strategy appears to be more limited that acquisition of the L2's phonotactic constraints. Thus, as described in Chapter 2, section 2.2.1, Cutler, Mehler, Norris and Seguí (1989, 1992) reported that English/French early, highly proficient bilinguals were only able to implement the segmentation strategy of their dominant language. Note however that Sanders, Neville and Woldorff (2002) found that L1 Japanese-L2 English and L1Spanish/L2 English late bilinguals implemented the rhythm-based segmentation strategy characteristic of their L2, i.e., a stress-based segmentation strategy.

The results from the present experiment suggest that bilinguals can acquire the frequency-based segmentation strategies characteristic of their two languages.

Summary of findings

The results obtained in this experiment can be summarized into the following conclusions:

(i) Acoustic-phonetic cues outrank statistical cues in the segmentation of the artificial language, as seen in the shift of segmentation preference found between experiments A and B.

(ii) The frequency-based segmentation of the artificial language obtained in Experiment A was not replicated.

(iii) The context modulates the segmentation preference of L1Basque bilinguals. Though statistically only marginally significant, the tendency obtained in this group replicates the context language effect obtained in the group of L1Basque bilinguals in Experiment A.

Chapter 5. Interaction between prosodic and frequency cues in adult speech segmentation

5.1 Introduction

In this chapter I discuss an artificial language learning experiment that examines the interplay of prosodic and statistical cues in the segmentation of fluent speech by Basque/Spanish bilinguals. Two are the aims of this experiment. First, I aim to investigate the relative weight given by adult speakers to prosodic as compared to statistical cues in speech segmentation. To that end a prosodic cue (i.e., changes in pitch) was introduced in the ambiguous language described in Experiment B, which hence contained also a frequency-based cue. It has been shown by Bion, Benavides-Varela and Nespor (2011) that changes in pitch lead to prominent-initial grouping of speech samples in adults and infants, following the Iambic/Trochaic Law (Hayes, 1995). The present experiment examines thus whether adult speakers segment the artificial language according to the segmentation signalled by the prosodic cue when this cue converges with the segmentation signalled by the frequency-based cue and when these two cues are in conflict.

Additionally, I aim to observe whether differences arise in the segmentation preferences of two groups of highly proficient bilingual speakers of a head-final, OV language (Basque) and a head-initial, VO language (Spanish) that differ in their first language (either Basque or Spanish). It has been proposed that pitch changes are the acoustic correlate of head-final structures (Nespor, Shukla, van de Vijver, Avesani, Schraudolf and Donati, 2008). This experiment aims thus to explore whether L1Spanish/L2Basque bilinguals implement the segmentation strategy characteristic of their L2, Basque (OV, head-final) or whether this group of bilinguals displays a different segmentation preference than L1Basque/L2Spanish bilinguals.

The previous chapters examined the frequency-based bootstrapping hypothesis in two artificial language learning experiments (Gervain, 2007). This hypothesis proposes that the frequency distribution and relative order of functors and content words in natural languages help infants set the basic configurational parameter of natural languages, i.e., the Head-Directionality parameter (Baker, 2001, Chomsky, 1981). The setting of this parameter provides the infant with an initial representation of the basic word order of the phrases of the language she is acquiring.

Suprasegmental information has been proposed as an alternative cue to the setting of the Head-Directionality parameter (Christophe, Guasti, Nespor, Dupoux and van Ooyen, 1997; Nespor, Guasti and Christophe, 1996; Nespor, Shukla, van de Vijver, Avesani, Schraudolf and Donati, 2008). Specifically, it has been argued that the relative prominence within phonological phrases is a perceptually available cue, which, due to its correlation with the relative order of heads and complements in natural languages (Nespor and Vogel, 1986), might be used by infants to set this parameter prior to the development of lexical items (Christophe et al., 1997; Christophe, Nespor, Guasti and van Ooyen, 2003; Nespor, 2001; Nespor, Guasti and Christophe, 1996; Nespor et al., 2008).

According to Nespor and Vogel (1986), the rightmost word within a phonological phrase receives the main prominence in right-branching, head-initial languages, whereas the leftmost word of a phonological phrase receives main prominence in left-branching, head-final languages. It has been proposed (Nespor et al., 2008) that the physical realization of prominence differs in head-initial vs. head-final languages and even in head-initial vs. head-final structures within a single language, in accordance with the Iambic/Trochaic Law (Hayes, 1995). Thus, increased duration characterizes prominence of the complement in head-initial structures, whereas higher pitch characterizes prominence of the complement in head-final structures (Nespor et al., 2008).

Table 16. Iambic/Trochaic Law

	Grouping	Example
Duration	Prominent-final: iambic	[– —][– —][– —][– —]
Pitch	Prominent-initial: trochaic	[x́ x] [x́ x] [x́ x] [x́ x]

Nespor et al.'s (1996, 2008) proposal that infants set the value of the Head-Directionality parameter on the basis of the main prominence of phonological phrases falls within the framework of the phonological bootstrapping hypothesis. This hypothesis claims that the speech signal contains available phonological cues that mark boundaries to syntactic constituents such as clauses and phrases (Gleitman and Wanner, 1982; Morgan and Demuth, 1996). These cues are exploited by the infants to conduct a first-pass analysis of the input, dividing it into smaller chunks (Gerken, Jusczyk and Mandel, 1994; Hirsh-Pasek et al, 1987; Jusczyk et al., 1992). This division enables infants to discover regularities that lead to the acquisition of certain aspects of the native language's syntax. As described in Chapter 2, section 2.2.2, though no one-to-one correlation exists between prosodic and syntactic constituents, intonational groups tend to align with major syntactic constituents (Cruttenden, 1986; Nespor and Vogel, 1986). Selkirk (1996, 2000) proposes the *Align XP* constraint, according to which, the right (or left) edge of a syntactic phrase (XP) must be aligned with the right (or left) edge of a phonological phrase.

Sensitivity to prosodic cues to phonological phrase boundaries has been attested in newborns. Christophe, Dupoux, Bertoncini and Mehler (1994), using the high-amplitude non-nutritive sucking procedure, found that French 3-day-old infants discriminated between phonemically identical CVCV stimuli that differed in the absence vs. presence of a phonological boundary between the two syllables. Stimuli were tokens of the sequence /*mati*/ spliced from either (a) a polysyllabic word (e.g., *mathé<u>mati</u>cien*), or (b) two adjacent words that belonged to different phonological phrases (e.g., *pyja<u>ma ti</u>ssé*). Infants that listened to a change between the two categories of stimuli (e.g., *mati* vs. *ma#ti*) showed higher sucking rates than the group of newborns that listened to the same category after the switch of stimulation (e.g., *mati* vs. *mati*). Christophe, Mehler and Sebastián-Gallés (2001) tested French newborns with similar stimuli, spliced from Spanish sentences (e.g., /*lati*/ spliced from *gelatina* vs. *Manuela timida*) and replicated the discrimination found with the French stimuli.

By 6- to 9-months of age infants are able to make use of the prosodic cues to syntactic phrases that correspond to phonological phrases in speech segmentation (Jusczyk, Hirsh-Pasek, Kemler Nelson, Kennedy, Woodward and Piwoz, 1992; Soderstrom, Seidl, Kemler Nelson and Jusczyk, 2003). Jusczyk et al. (1992) found that 9-month-old infants listened longer to sequences in which 1 second pauses were inserted at the subject-verb phrase boundary (e.g., *Did you/spill your cereal?*) than to sequences in which the pause was inserted within a phrase (e.g., *Did you spill/your cereal?*). Similarly, Soderstrom, Seidl, Kemler Nelson and Jusczyk (2003) familiarized English-learning 6- and 9-month-olds with phonemically identical sequences that formed either a well-formed syntactic phrase (e.g., *new watches for men*), or a sequence that spanned a subject-verb phase boundary (e.g., *gnu#watches for men*). During test phase, infants listened to passages that contained the two sequences (e.g., *At the discount store, new watches for men are simple and stylish*. […] vs. *In the field, the old frightened gnu # watches for men and women seeking trophies*. […]). Both the 6- and 9-month-olds listened longer to the passages containing the familiarized well-formed phrase than to the passages containing the familiarized but ill-formed sequence.

This sensitivity to the prosodic markers of major syntactic phrase boundaries is nonetheless limited, due to the above mentioned lack of total isomorphism between syntactic and prosodic constituents. This limitation is illustrated by the results in Gerken, Jusczyk and Mandel (1994), who replicated the sensitivity to prosodic cues that signal the subject-verb phrase boundary in 9-month-old infants, but found that this sensitivity was restricted to NP subjects (e.g., *The caterpillar wasn't hungry anymore.*) and did not extend to pronoun subjects (e.g., *He wasn't hungry anymore.*). In a similar experiment to Jusczyk et al. (1992), 9-month-old infants showed no differences in listening times between sentences that contained a pause between the pronoun subject and the verb (e.g., *He/wasn't hungry anymore*) and sentences in which the pause was inserted after the verb (e.g., *He wasn't/hungry anymore*). Pronoun subjects, unlike full NP subjects, do not form a phonological phrase on their own, but are incorporated to the phonological phrase that contains the verb (Selkirk, 1984, 1996).

Phonological phrase boundaries have also been shown to constrain word segmentation and lexical access in infants and adults (Christophe, Peperkamp, Pallier, Block and Mehler, 2004; Gout, Christophe and Morgan, 2004). Gout, Christophe and Morgan (2004) trained 10- and 13-month-old English-learning infants to turn their head whenever they heard a target bisyllable (e.g., *paper*) and subsequently presented the infants with sentences that contained the same phonemic sequence, either embedded in a phonological phrase (e.g., [*The college*] [*with the biggest <u>paper</u> forms*][*is best*]),[33] or straddling a phonological phrase boundary (e.g., [*The butler*] [*with the highest <u>pay</u>*][*<u>per</u>forms the most*]). Both the 10- and the 13-month-olds turned their heads more often upon hearing the sentences in which the target was contained within a phonological phrase, than upon hearing the sentences in which the target spanned a phonological boundary. Hence, the presence of a phonological phrase boundary prevented infants from taking the two adjacent syllables as a word candidate. Similarly, Christophe et al. (2004), in a word-monitoring task, found that activation of lexical competitors (e.g., *chagrin*) to a target word (e.g., *chat*) was constrained when the competitor straddled a phonological phrase boundary (e.g., [*D'après ma sœr*], [*le gros <u>chat</u>*] [*<u>grim</u>pait aux arbres*]...), but not when the competitor straddled a prosodic word boundary within a single phonological phrase (e.g., [*Le livre*] [*racontait l'histoire*] [*d'un <u>chat grin</u>cheux*]...).

These results suggest that infants are sensitive to the acoustic markers of phonological phrase boundaries from a very early age and make use of these markers to segment the input prior to having lexical knowledge. As mentioned above, the prosodic correlate to the Head-Directionality parameter, i.e., the location of the main prominence within phonological phrases, has been proposed as a potential cue to bootstrapping this parameter (Christophe et al., 1997; Nespor, Guasti, Christophe, 1996; Nespor et al., 2008). Christophe, Nespor, Guasti and van Ooyen (2003) provided evidence on both adults' and infants' ability to discriminate a head-initial language (French) from a head-final language (Turkish) based on the location of the main prominence within phonological phrases.

Christophe et al. (2003) conducted a categorization task with adult French listeners and a discriminability task using the high-amplitude non-nutritive sucking procedure with French-learning 2- to 3-month-olds. Stimuli were sentences in French (head-initial language) and Turkish (head-final), matched in number of syllables and the

[33] Where the brackets represent the domains of a phonological phrase.

location of word boundaries, stresses and boundaries of intonational and phonological phrases (e.g., French: [*Le grand orang-outáng*] [*etait enervé*]; Turkish: [*Yení kitabɪmɪ*] [*almák istiyor*]).[34] Although French and Turkish have opposite values of the Head-Directionality parameter, the two languages share a series of prosodic properties (i.e., word-final stress, a similar syllabic structure and no vowel reduction), ruling hence these properties out as potential cues to the discrimination of the French and Turkish stimuli. The original sentences were resynthesized with a series of Dutch phonemes, hence removing phonetic information specific of the two languages. The resynthesized sentences differed only in the location of the prominence within phonological phrases. Prominence fell on the rightmost word of the phonological phrase in the original French (head-initial) sentences (e.g., [*leplem pelemepém*] [*epe pemelsé*]), and on the leftmost word of the phonological phrase in the original Turkish (head-final) sentences (e.g., [*jemé pepepeme*] [*elmép espejel*]).

In the categorization task, participants were instructed to sort the stimuli into two categories. Four of the five adult participants were able to categorize the stimuli better than chance (65%). In the discriminability task, infants listened to an experimental switch, i.e., a change from the originally French sentences to the originally Turkish sentences (or vice versa) and a control switch, i.e., a change between sentences that were originally from the same language. The 2- to 3-month-olds showed higher sucking rates in the experimental switch than in the control switch.

The results of these experiments show that both adults and infants are sensitive to the relative prominence within phonological phrases and are able to make use of the relative location of this prominence to discriminate between head-initial and head-final languages. Nespor, Shukla, van de Vijver, Avesani, Schraudolf and Donati (2008) examined the physical realization of the prominence of heads and complements within phonological phrases, in head-initial and head-final languages and proposed that infants might set the value of the Head-Directionality parameter relying on two laws: the Complement Law (Cinque, 1993; Nespor et al., 2008) and the Iambic/Trochaic Law (Hayes, 1995).

According to the Complement Law, in the broad focus intonation characteristic of declaratives, complements receive main prominence, regardless of their location relative to heads (Nespor et al., 2008:4). Cinque (1993) proposes that prominence is a

[34] The brackets signaling the domain of the phonological phrase and the diacritic signaling the main prominence.

reflection of depth of embedding of the constituents. Complements are more deeply embedded (i.e., complements are dominated by more projections) than heads and hence receive main prominence.

The Iambic/Trochaic Law states that elements that contrast in intensity form groupings with initial prominence whereas elements that contrast in duration form groupings with final prominence (Hayes, 1995:80). Hay and Diehl (2007) propose that this law results from a general auditory bias. Thus, Hay and Diehl (2007) examined the perceptual grouping of sequences of both linguistic (i.e., synthesized tokens of the syllable /*ga*/) and non-linguistic stimuli (i.e., square wave segments) with speakers of English and French. Listeners were presented with sequences that contained alternating stimuli that differed in either duration or intensity. Participants' task consisted on grouping the stimuli into either a strong-weak rhythmic pattern (i.e., a trochaic pattern), or a weak-strong rhythmic pattern (i.e., an iambic pattern). Both the French and English listeners grouped the sequences that contained elements with different durations into an iambic pattern (i.e., with the prominent element in final position) more often than control stimuli that had uniform duration. On the other hand, participants grouped the stimuli that contained variations in intensity into a trochaic pattern (i.e., with the prominent element in initial position) more often than monotonous control stimuli. This grouping was found both in linguistic and non-linguistic stimuli. Additionally, Cooper and Meyer (1960) showed that rhythmic structure in music followed the predictions of this law. Last, Peña, Bion and Nespor (2011) found that participants appeared to form groupings of visual stimuli according to a visual correlate of the Iambic/Trochaic law.

Nespor et al. (2008) analyzed the location and acoustic realization of prominence in sequences of heads and complements embedded in sentences, produced by speakers of French (head-initial), Turkish (head-final) and German (mainly head-initial, though head-final in subordinate clauses). The French and Turkish stimuli shared number of syllables and stress patterns, and contained lexical items that have similar pronunciation in both languages (e.g., French: *En français antique kilim se dit,* i.e., *In French ancient kilim is said.* vs. Turkish: *Türkçe'de kilim için denir,* i.e., *In Turkish for the kilim is said.*). In German, stimuli consisted of pairs of sentences, both members of the pair being identical except for the order of a head (i.e., a verb) and a complement (i.e., the direct object of the verb) (e.g., *Der Abend wird gut werden, weil ich Papa sehe.* vs. *Der Abend wird gut werden, denn ich sehe Papa.*, i.e., *The evening will be pleasant, because I (will) see Papa.*). Analysis of the Turkish and French stimuli showed that main prominence fell on the vowel of the stressed syllable of the complement in both languages as predicted by the Complement Law. In the Turkish stimuli, the vowel of the stressed syllable of the complement had higher intensity and pitch and was longer than the vowel of the stressed syllable of the head. In French on the other hand, the vowel of the stressed syllable of the complement had higher intensity and was longer than the vowel of the stressed syllable of the head. Analysis of the German stimuli revealed that, when the complement (the direct object of the verb) appeared at initial position, its stressed syllable had higher pitch and intensity, whereas the stressed syllable of the complement was characterized by longer duration when the complement followed the head.

These results show that the realization of the main prominence in sequences of heads and complements appears to follow the Iambic/Trochaic Law, both cross-linguistically and within a language. Prominence is thus realized through increased duration in head-initial structures and through higher pitch (and intensity) in head-final structures. Nespor et al. (2008) suggest that this different realization of prominence in head-final vs. head-initial structures could enable infants to set the value of the Head-Directionality parameter, without prior location of phonological phrase boundaries.

Bion, Benavides-Varela and Nespor (2011) examined whether these two acoustic markers of prominence (i.e., pitch and duration) influenced the grouping of speech samples as predicted by the Iambic/Trochaic Law. To that end, Bion et al. (2011) conducted two artificial language learning experiments with adult speakers of Italian and Italian-learning 7-month-olds. Adults were familiarized in sequences of ten CV syllables that either: (a) alternated in pitch, (b) alternated in duration, or (c) had constant duration and pitch values. Test stimuli consisted on pairs of CV syllables with constant duration and pitch that (a) had occurred adjacently in the familiarization with prominence in the first syllable, (b) had occurred adjacently in the familiarization with prominence in the second syllable, or (c) did not occur adjacently in the familiarization. The task consisted on judging for each pair of syllables whether the pair had occurred adjacently in familiarization. Participants familiarized with sequences of alternating pitch remembered better the pairs of syllables that occurred in familiarization with prominence on the first syllable of the pair, than pairs of syllables that occurred with prominence in the second syllable. The opposite pattern was found in the responses of the participants familiarized with sequences of alternating duration, i.e., participants were better at remembering the pairs that had occurred with prominence in the second syllable. Adult listeners grouped hence the stimuli in accordance to the predictions of the Iambic/Trochaic Law, i.e., changes in pitch led to prominence-initial groupings of the stimuli, whereas changes in duration led to prominence-final groupings of the stimuli.

Bion et al. (2011) investigated whether these acoustic markers similarly influenced the grouping of stimuli in 7-month-old infants, by means of the HPP. Infants were familiarized in sequences of six CV syllables that alternated in pitch, duration, or had a constant duration and pitch. In the test phase, infants listened to repetitions of pairs of syllables with constant duration and pitch, but which had (a) occurred adjacently in the familiarization with prominence in the first syllable, or (b) occurred adjacently in the familiarization with prominence in the second syllable. The group of infants familiarized with the sequences of alternating pitch preferred the test pairs which, in familiarization, had received prominence on the first syllable, i.e., infants looked longer to these pairs than to the pairs that occurred in familiarization with final prominence. The group of infants familiarized with alternating duration, however, did not show any preference. Bion et al. (2011) concluded that the bias to group stimuli that differ in pitch into prominence-initial units (i.e., trochees) appears to emerge at an earlier point in development than the bias to group stimuli that differ in duration into prominence-final units (i.e., iambs).

Indeed, research on perception grouping by duration with both infants and adults has revealed that duration-based grouping also appears to be influenced by language experience (Iversen, Patel and Ohgushi, 2008; Yoshida, Iversen, Patel, Mazuka, Nito, Gervain and Werker, 2010). Thus, Iversen et al. (2008) showed that English speakers preferred short-long groupings of sequences of tones that contained changes in duration, as predicted by the Iambic/Trochaic Law. However, Japanese speakers mostly chose a long-short grouping of the stimuli. Iversen et al. (2008) suggest that this asymmetry might result from differences in the order of functors and content words in natural languages, given that functors tend to be shorter than content words. Consequently, speakers of a head-final (hence functor final) language such as Japanese might be biased towards a long-short grouping of stimuli that contain changes in duration, as opposed to head-initial languages such as English.

5.2 The interplay of prosodic and statistical cues in adult speech segmentation

As mentioned in the introduction, this experiment investigates how prosodic and frequency-based (i.e., statistical) cues interact in adult speech segmentation and specifically examines the segmentation preferences of Basque/Spanish bilinguals. The artificial language in the present experiment shared lexicon and structure with the artificial language described in Experiment B, but contained an added prosodic cue, i.e., the infrequent CV elements were synthesized with a f0 of 120 Hz, whereas frequent CV elements were synthesized with a f0 of 100 Hz. Hence, pitch rises and falls alternated in the artificial language.

If, as predicted by Bion et al. (2011), elements that contrast in pitch height form groupings with initial prominence (i.e., form a trochaic pattern), a general frequent-final segmentation preference is expected to obtain. Furthermore, if, as proposed, the Iambic/Trochaic Law is a general auditory bias (Hay and Diehl, 2007), this segmentation preference is expected to obtain regardless of the value of the Head-Directionality parameter in the participants' native language. Hence, a general frequent-final segmentation preference is expected to obtain in the two groups of participants tested, i.e., L1Basque/L2Spanish and L1Spanish/L2Basque bilinguals (Basque: OV, Spanish VO).

Additionally, a difference in the segmentation preferences of L1Basque/L2Spanish and L1Spanish/L2Basque bilinguals is predicted, due to the integration of the information provided by this prosodic cue and the statistical cue present in the input, i.e., the frequency distribution of frequent and infrequent elements discussed in experiments A and B. The statistical information cues L1Basque bilinguals toward a frequent-final segmentation of the ambiguous language, which converges with the segmentation cued by prosodic information. However, statistical information cues L1Spanish bilinguals toward a frequent-initial segmentation of the language, in conflict with the frequent-final segmentation cued by prosody.

Thus, the frequency-based bootstrapping hypothesis proposes that adult speakers build a representation of the frequency distribution and relative order of functors (frequent elements) and content words (infrequent elements) characteristic of the native language (Gervain, 2007; Gervain et al., submitted). As shown by Gervain (2007) and replicated by the results from experiment A in Chapter 3, this representation influences the segmentation preference of native speakers of VO and OV languages. In Experiment A, speakers of an OV language (i.e., Basque) chose frequent-final segmentation of an ambiguous artificial language more often than speakers of VO languages (Spanish, French and English).[35] However, Experiment B failed to replicate the frequent-final segmentation preference found in native speakers of Basque both in Gervain (2007) and in Experiment A.

I argued that acoustic-phonetic properties of the input might have overridden the statistical cue provided by the frequency distribution of the elements in the artificial language. As mentioned above, in the present experiment, both the prosodic cue (i.e., pitch changes) and the statistical cue (i.e., the frequency distribution of the elements) provide cues supporting a preference for a frequent-final segmentation to speakers of OV languages. Integration of these two cues might thus override the bias created by acoustic-phonetic properties of the input of Experiment B and result in an increased frequent-final segmentation preference in the group of L1Basque bilinguals. Regarding speakers of VO languages, the prosodic cue signals a frequent-final segmentation of the input, whereas the statistical cue supports a frequent-initial segmentation. Therefore, the group of L1Spanish bilinguals tested, upon hearing these two conflicting cues, might not choose the frequent-final segmentation preference predicted by the prosodic cue as often as the group of L1Basque bilinguals.

Alternatively, I proposed that the general frequent-initial segmentation preference obtained in Experiment B might result from a universal preference for frequent-initial segmentation. The addition of the prosodic cue in Experiment C, signalling a frequent-final segmentation, might override this default bias and result in greater frequent-final segmentation in both groups of participants. However, it is also posssible that only the convergent information provided by this prosodic cue and the statistical information that cues a frequent-final segmentation to the group of L1Basque bilinguals suffices to override the universally preferred strategy. This would entail a bigger frequent-final segmentation in the group of L1Basque bilinguals, and a smaller or even no frequent-final segmentation in L1Spanish bilinguals.

[35] Note however that the difference between Basque natives and English and Spanish natives was only marginally significant, probably due to acoustic-phonetic properties of the German voice used to synthesize the stimuli.

5.2.1 Methodology

5.2.1.1 Participants

31 undergraduate students (23 females) from the University of the Basque Country (UPV/EHU), with a mean age of 21.9 years (age range 18 to 37) took part in this experiment. Participants were sorted into two groups, depending on their native language (either Basque or Spanish). The group of L1Basque bilinguals was addressed and received the instructions in Basque (16 participants, mean age 20.94, age range 18 to 29). The group of L1Spanish bilinguals, on the other hand, was addressed and received the instructions in Spanish (15 participants, mean age 22.93, age range 18 to 37).

As in the previous experiments, the linguistic background of the participants was collected by means of a questionnaire, which revealed that both groups of bilinguals acquired their L2 at around 3 to 4 years of age. The L1Spanish/L2Basque participants had completed or were conducting studies in Basque at the UPV/EHU at the time the experiment was conducted, possessed a degree in Basque studies, or had the EGA (Basque Language Certificate, C1 level in the Common European Framework). Participants were naïve to the purpose of the experiment and received a 4€ compensation for their participation.

5.2.1.2 Materials

The artificial language was the same as in Experiment B, with one added manipulation, i.e., a 20 Hz f0 rise in the infrequent syllables. Thus, the basic unit of this artificial language was a six-syllable long basic unit with the structure: *aXbYcZ*. *a, b,* and *c* represented frequent categories that consisted of a CV token each (*a* = *fi*; *b* = *nu*; *c* = *pe*), whereas *X, Y* and *Z* represented infrequent categories that consisted of 9 CV tokens each (e.g., *X* = *fe, ka; Y* = *li, sa; Z* = *po, tu*). Hence, frequent categories were 9 times more frequent than infrequent categories. The lexicon of CV syllables was the exact same lexicon used in Experiment B. The basic unit was concatenated 377 times creating a 9 minute and 3 second long stream. Intensity increased gradually during the first minute and decreased gradually during the last minute of the stream in order to ensure ambiguity.

The CV tokens from the frequent categories *a*, *b* and *c* were synthesized with 100 Hz and had a constant duration of 120 msec per phoneme. The CV tokens from the infrequent categories were similarly synthesized with a constant duration of 120 msec per phoneme, but with a f0 of 120 Hz, i.e., 20 Hz higher than frequent syllables. Hence, strict alternation of frequent and infrequent syllables tallied with strict alternation of pitch falls and rises.

Table 17. Experiment C. Familiarization

Segmentation preference	**Structure**	**Familiarization**
	…**a**X**b**Y**c**Z**a**X**b**Y**c**Z**a**X**b**Y**c**Z **a**X**b**Y**c**Z**a**X**b**Y**c**Z**a**X**b**Y**c**Z…	…fi**né**nu**lí**pe**kí**fi**ká**nu**ló**pe**lé**fi**ká**nu**lí**p e**pí**fi**pá**nu**fú**pe**kú**fi**né**nu**ná**pe**pí**fi…
Frequent-final	…**a[***X**b**Y**c**Z**a***]**X**b**Y**c**Z**a**X**b**Y…	…fi**[*né*nu*lí*pe*kí*fi]ká**nu**ló**pe**lé**fi**ká**…
Frequent-initial	…**a**X**[*b*YcZaX]b**Y**c**Z**a**X**b**Y…	…fi**né[**nu*lí*pe*kí*fi*ká*]nu**ló**pe**lé**fi**ká**…

Test stimuli consisted of 36 six-syllable long items, with the same phoneme sequences as in Experiment B. However, all infrequent syllables of the present stimuli were, as in the familiarization, synthesized with a f0 of 120 Hz. 18 of the items started thus with a prominent infrequent syllable, instantiating a frequent-final order (e.g., *XbYcZa*: *KÁnuLÍpePÓfi*). The remaining 18 items ended with a prominent infrequent syllable (e.g., *aXbYcZ*: *fiLÚnuSÁpeKÍ*), displaying a frequent-initial order.

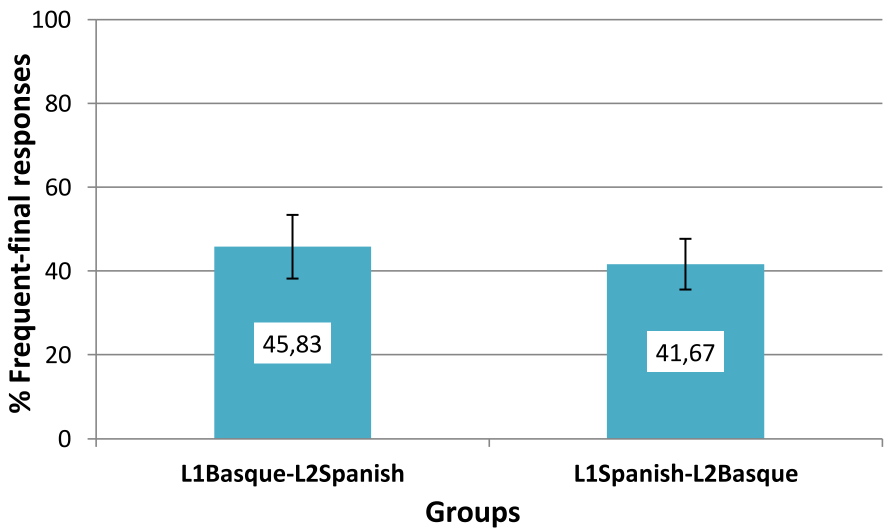
5.2.1.3 Procedure

As in Experiment B, participants were tested individually in a quiet room, in the psycholinguistics laboratory at the University of the Basque Country (UPV/EHU). Procedure was similar to experiments A and B.

5.2.2 Results

Data analysis was conducted using SPSS software. Kolmogórov-Smirnov tests revealed normal distribution of the dependent variable in both groups. One-sample t-tests were conducted for each group, in which the mean score of the sample was compared to a test value of 18 (out of 36, a value of 18 would amount to chance). Both the group of L1Basque bilinguals (M = 16.5, SD = 10.94) and the group of L1Spanish bilinguals (M = 15, SD = 8.44) chose a frequent-initial segmentation more often than a frequent-final segmentation. However, this preference did not significantly differ from chance in neither of the groups (L1Bsq: t(15) = -.548, p = .592; L1Sp: t(14) = -1.378, p = .190). Pair-wise comparison of the two groups by means of independent t-tests revealed no significant differences (t(29) = .425, p = .674).

Graph 14. Experiment C. Percentage of frequent-final responses per group



The bias toward a frequent-final segmentation preference predicted by the the Iambic/Trochaic Law (Bion et al., 2011; Hayes, 1995; Nespor et al., 2008) was not found in the results of the bilingual participants. Also, the potential difference in the segmentation preference between the groups of L1Basque and L1Spanish participants that might have emerged due to the joint influence of the two cues present in the input (i.e., prosodic cue: pitch changes, and statistical cue: frequency distribution of frequent and infrequent elements) was not observed either.

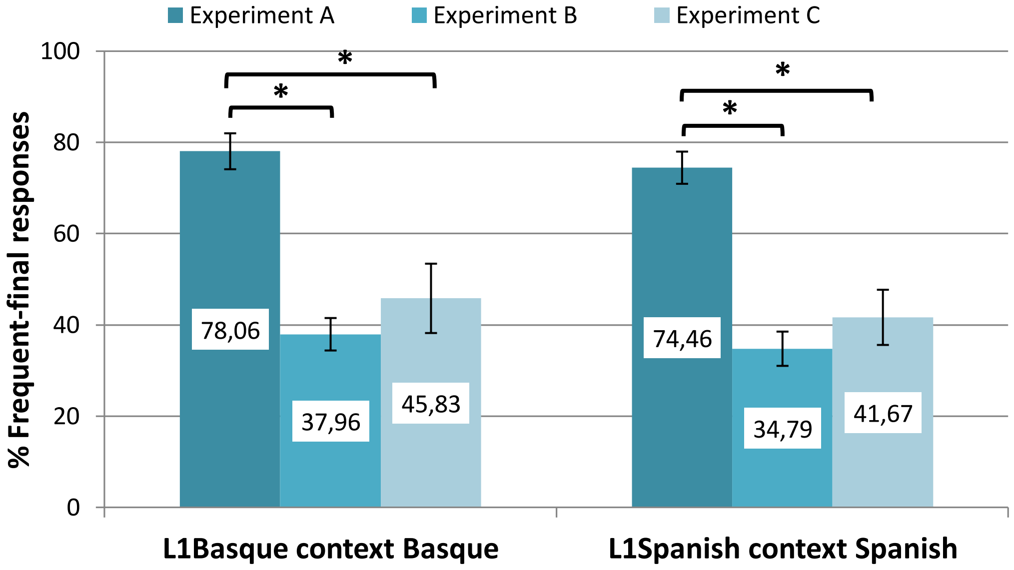
5.2.2.1 Experiment A vs. B vs. C

The segmentation preferences of the two groups tested in the present experiment was compared to the segmentation preferences of similar groups in experiments A and B. The group of L1Basque bilinguals in the present experiment received the instructions in Basque and is hence similar to the group of L1Basque context Basque bilinguals tested in experiments A and B. The group of L1Spanish bilinguals, on the other hand, received the instructions in Spanish, as did the group of L1Spanish context Spanish bilinguals in experiments A and B.

An ANOVA with the factor *experiment* was conducted with the results of the L1Spanish participants across experiments A, B and C, and a significant main effect was obtained ($F(2.64) = 29.199$, $p < .001$). Pair-wise comparisons were conducted by means of independent t-tests. The t-tests revealed a significant difference in the segmentation preference of the L1Spanish context Spanish group between experiments A ($M = 26.81$, $SD = 7.07$) and C ($M = 15$, $SD = 8.44$) ($t(44) = 4.983$, $p < .001$) and A ($M = 26.81$, $SD = 7.07$) and B (12.52, $SD = 6.19$) ($t(50) = 7.503$, $p < .001$). The segmentation preference of the L1Spanish context Spanish group in experiments B ($M = 12.52$, $SD = 6.19$) and C ($M = 15$, $SD = 8.44$) did not differ significantly ($t(34) = -1.017$, $p = .316$).

Similarly, an ANOVA with the factor *experiment* was conducted with the results of the L1Basque participants across experiments A, B and C, and a significant main effect was obtained ($F(2.60) = 20.401$, $p < .001$). Again, pair-wise comparisons were conducted by means of independent t-tests. The t-tests revealed a significant difference in the segmentation preference of the L1Basque context Basque group between experiments A ($M = 28.10$, $SD = 6.34$) and C ($M = 16.5$, $SD = 10.94$) ($t(34) = 3.986$, $p < .001$) and experiments A ($M = 28.10$, $SD = 6.34$) and B ($M = 13.67$, $SD = 6.66$) ($t(45) = 7.498$, $p < .001$). The segmentation preference of the L1Basque context Basque group in experiments B ($M = 13.67$, $SD = 6.66$) and C ($M = 15$, $SD = 8.44$) did not differ significantly ($t(41) =-1.059$, $p = .296$).

Graph 15. Comparison of frequent-final responses in experiments A, B and C



5.3 Discussion

The present chapter aimed to examine whether the presence of the prosodic correlate of head-final (OV) languages, namely, changes in pitch, biased the listeners' segmentation preferences of Experiment B's ambiguous artificial language toward a prominent-initial grouping of the stimuli (Bion et al., 2011; Nespor et al., 2008). This predicted bias was not found in the segmentation preferences of neither the groups of L1Basque/L2Spanish or L1Spanish/L2Basque participants.

It should be pointed out that, unlike Bion et al. (2011), which presented participants with prosodically flat stimuli during the test phase, in the present experiment the test stimuli consisted of pairs of prominent-initial and prominent-final stimuli, therefore containing the prosodic cue. It is hence possible that participants in Experiment C chose the preferred test item based on the prosodic cue present in the test items rather than chosing based only on the familiarization. A control experiment with prosodically flat test stimuli is pending, in order to rule out this possibility.

In Experiment B, all groups of participants displayed a frequent-initial segmentation preference of the artificial language. I argued that acoustic-phonetic properties of the input might have caused a bias toward frequent-initial segmentation. In the present experiment the prosodic cue (i.e., pitch changes) should signal the opposite segmentation, namely a frequent-final segmentation of the language. Infrequent syllables had higher pitch (hence more prosodic prominence), and according to the Iambic/Trochaic Law (Hay and Diehl, 2005; Hayes, 1995) and to Bion et al's (2011) findings, this fact should lead to a trochaic-rhythm parse, with infrequent syllables in initial position and frequent syllables in final position. The frequent-initial segmentation preference found in the present experiment suggests that the prosodic cue provided by pitch changes did not override the bias created by the acoustic-phonetic properties of the input, as predicted by Mattys et al.'s (2005) hierarchy of segmentation cues.

According to Mattys, White and Melhorn's (2005) hierarchical approach to speech segmentation, presented in Chapter 2, section 2.3 of the present dissertation, the available cues to speech segmentation are sorted into a three-ranked structure, according to the relative weight that listeners confer to each of these cues. Thus, speech segmentation in adults is driven by lexical cues (tier I). In the absence or ambiguity of lexical information, listeners rely on segmental cues (tier II), such as phonotactics and acoustic-phonetic cues (e.g., allophonic variation, coarticulation). Prosodic cues are proposed as the lowest ranked cue (tier III) that guides segmentation whenever lexical and segmental cues are impoverished or ambiguous. Thus, adult listeners have been shown to rely on rhythmic cues only when encountering a degraded signal (Mattys, 2004; Mattys et al., 2005).

Mattys et al. (2005) conducted a series of lexical decision tasks using the cross-modal priming fragment paradigm with adult speakers of English, in which rhythmic cues (i.e., stress) were pitted against acoustic-phonetic (i.e., coarticulation), phonotactic and lexical cues. In two experiments, participants listened to nonsense utterances that contained SW (i.e., a strong syllable followed by a weak syllable) and WS primes containing either coarticulatory or phonotactic cues that favoured or disfavoured the segmentation of the prime from the surrounding context. English is a stress-timed language, characterized by the opposition of strong and weak syllables (Cutler et al., 1983, 1986). Speakers of stress-timed languages have been shown to segment the input at the onset of strong syllables (Cutler and Norris, 1988). Hence, prosody would cue segmentation of SW primes, but would not cue segmentation of WS primes. A parallel set of utterances was created, in which primes were replaced with distorted speech (i.e., by superimposition of SW and WS primes). The results of these experiments revealed that, in the non-distorted stimuli, priming was facilitated by favourable acoustic-phonetic (i.e., coarticulation) or phonotactic cues, regardless of the cue provided by stress. The opposite pattern was found in the stimuli with distorted speech, i.e., in which coarticulatory or phonotactic information was unavailable. Thus, priming was facilitated in SW stimuli, as predicted for speakers of stress-timed languages.

In a subsequent experiment, Mattys et al. (2005) pitted lexical cues against prosodic cues. SW and WS primes were embedded in utterances, preceded by either a real English word favouring segmentation of the prime, or a nonsense word which did not favour segmentation. Again, a parallel set of utterances was created, in which the whole utterance was replaced with distorted speech (i.e., by superimposition of test utterances). In the non-distorted stimuli, priming was facilitated by the primes preceded by a real English word. In the distorted stimuli, however, priming was facilitated by stress, i.e., in SW stimuli.

Therefore, prosodic cues are available speech segmentation cues to adult listeners, albeit overridden by acoustic-phonetic, phonotactic and lexical cues. Whenever factors (e.g., signal quality) render acoustic-phonetic, phonotactic and lexical information unintelligible, adult listeners make use of prosodic cues such as rhythm to segment speech.

A hypothesis was suggested for a potential difference in the segmentation preferences of L1Basque and L1Spanish participants, as a result of the potential integration of the two cues present in the stimuli, i.e., prosodic and statistical information. Following the predictions of the frequency-based bootstrapping hypothesis, the relative frequency of frequent elements in the artificial language would cue L1Basque participants (Basque: OV) to a frequent-final segmentation of the stream, converging with the segmentation preference cued by prosodic information. On the other hand, the relative frequency of the artificial language's frequent elements would cue L1Spanish participants (Spanish: VO) to a frequent-initial segmentation, conflicting with the frequent-final segmentation cued by prosodic information. Integration of these two cues (i.e., statistical and prosodic) might have thus resulted in a greater frequent-final segmentation preference in the group of L1Basque bilinguals. However, no such difference in segmentation preference of the two groups was observed, which suggests that listeners might not have integrated the information of these two cues.

A possibility that needs to be considered is that these two cues were integrated but were outranked by the acoustic-phonetic information. It is possible that the integration of statistical and prosodic cues does not suffice to outweigh a cue ranked higher in the hierarchical structure of segmentation cues, tallying with Mattys et al.'s (2005) finding that adult speakers weigh more heavily lexical cues (tier I) than convergent coarticulatory and phonotactic cues (tier II). As described in Chapter 2, section 2.3, both statistical and prosodic cues have been shown to be outweighed by acoustic-phonetic information (Fernandes, Ventura and Kolinsky, 2007; Mattys, White and Melhorn, 2005).

As mentioned in section 5.2 of the present chapter, I proposed an alternative explanation for the general frequent-initial segmentation preference obtained in Experiment B, namely, that this strategy might be the unmarked or default strategy. Would this segmentation be the universally preferred segmentation, the absence of a frequent-final segmentation in Experiment C would entail that neither the prosodic cue, nor the convergent cues provided by prosodic and statistical information to the group of L1Basque bilinguals sufficed to override this default preference.

As opposed to the adult listeners' top-down approach of the hierarchical structure of segmentation cues, infants' speech segmentation abilities have been shown to follow a bottom-up approach. Statistical and prosodic information are the two earliest available segmentation cues. As described in Chapter 2, infants are able to segment words from speech based on transitional probabilities and rhythmic cues at around 7 months of age (Johnson and Tyler, 2010; Jusczyk, Houston and Newsome, 1999). Phonotactic and acoustic-phonetic cues (e.g., allophonic variation, coarticulation) require more detailed knowledge of the properties of the language under acquisition and have been found to emerge around 8 to 10½ months of age (Curtin, Mintz and Byrd, 2001; Friederici and Wessels, 1993; Jusczyk, Hohne and Bauman, 1999; Mattys and Jusczyk, 2001). As a result of this gradual tuning to the characteristics of the language under acquisition, infants rely primarily on statistical and prosodic cues to segment the input. Indeed, 9-month-old infants have been shown to weigh prosodic cues more heavily than probabilistic phonotactic cues (Mattys, Jusczyk, Luce and Morgan, 1999). Hence, a frequent-final segmentation preference of the artificial language would be expected in an experiment that tested infants.

Regarding the potential differences between L1Spanish and L1Basque infants, cue integration has been found in 8- to 9-month-olds. Morgan (1994) and Morgan and Saffran (1995) showed that convergent statistical and rhythmic cues enhanced segmentation of word-like items. Hence, it is possible that L1Basque infants would benefit from the coincident information provided by both the relative frequency of the artificial language's elements and the pitch changes, resulting in greater frequent-final segmentation than L1Spanish infants.

Summary of findings

The results obtained in this experiment can be summarized into the following conclusions:

(i) The presence of the prosodic correlate (i.e., pitch changes) to the head-final value of the Head-Directionality parameter does not induce an increase in prominence-initial segmentation of the ambiguous artificial language by adults.

(ii). Acoustic-phonetic properties of the input might have overridden the prosodic cue provided by changes in pitch, determining the segmentation of the artificial language, as predicted by Mattys, White and Melhorn's (2005) hierarchical account of speech segmentation. Alternatively, the prosodic cue might have not sufficed to overcome a default frequent-initial segmentation preference.

(iii.) The convergent information provided by the statistical and prosodic cues does not outweigh the bias created by either acoustic-phonetic information or the preference for the unmarked frequent-initial segmentation strategy.

Chapter 6. General discussion and conclusions

In Chapters 3, 4 and 5, I presented and discussed three artificial language learning experiments (Experiments A, B and C) that investigated adult monolingual and bilingual speakers' abilities to make use of statistical and prosodic cues to segment the input into chunks.

The main aim of these experiments was to examine whether bilingual speakers were able to implement the segmentation strategies of their two languages, or whether they exclusively used the strategies characteristic of their L1 in speech segmentation. Additionally, the experiments of the present dissertation aimed to explore the relative weight given to statistical and prosodic cues by adult speakers, within the framework of the hierarchical account of segmentation cues proposed by Mattys, White and Melhorn (2005). Last, in these experiments I sought to observe whether adult speakers make use of a type of statistical information and a type of prosodic information which have been proposed as potentially crucial to the bootstrapping of syntax acquisition in infants, namely the frequency distribution of frequent elements and their order relative to infrequent elements, mirroring functors and content words in natural languages, and the acoustic realization and relative location of prominence within phonological phrases. It has been proposed that these two cues correlate with the order of heads and complements in natural languages, and might hence enable infants to build a rudimentary representation of word order, setting the value of the major configurational parameter, i.e., the Head-Directionality parameter. These three topics will be separately addressed in what follows.

6.1 Bilinguals' implementation of segmentation procedures

As mentioned above, the main aim of this dissertation was to explore the segmentation abilities of bilingual speakers in fluent speech. Experiments A and B (Chapters 3 and 4) focused on a statistical cue to the segmentation of the input, namely frequency. As presented in Chapter 2, section 2.1, both infants and adults are extremely sensitive to the statistical information contained in the input, in the sense that they use it not only to locate word- and phrase- boundaries, but also to extract regularities (Gómez and Maye, 2005; Saffran, Aslin and Newport, 1996; Saffran, Newport and Aslin, 1996; Peña, Bonatti, Nespor and Mehler, 2002).

In experiments A and B, I presented Basque/Spanish bilingual speakers with a prosodically flat ambiguous artificial language characterized by a strict alternation of frequent and infrequent syllables. According to the predictions of the frequency-based bootstrapping hypothesis (Gervain, 2007)—which will be addressed later in the discussion—, a bigger tendency for frequent-final segmentation (i.e., segmentation in which frequent elements occur at final position) of the ambiguous language was expected to obtain in speakers of head-final, OV languages as compared to speakers of head-initial, VO languages.

The results of experiments A (Chapter 3) and B (Chapter 4) revealed that the L1Basque/L2Spanish bilingual speakers' segmentation preferences were modulated by the context language (i.e., the language in which participants received the instructions and were addressed) in the direction predicted by the frequency-based bootstrapping hypothesis. Thus, L1Basque bilinguals which received the instructions in Basque chose a frequent-final segmentation of the language more often than L1Basque bilinguals addressed in Spanish. However, the context language did not exert such an influence in the L1Spanish/L2Basque bilinguals' segmentation preferences of the artificial language.

These results suggest that L1Basque bilinguals, unlike L1Spanish bilinguals, were indeed able to make use of the segmentation strategies of both their L1 (Basque) and their L2 (Spanish). Age of acquisition of the L2 has been one of the factors shown to influence L2 processing (Weber-Fox and Neville, 1996), but it cannot account for the asymmetry found between the segmentation abilities of L1Basque and L1Spanish bilinguals, given that both groups of bilinguals acquired their respective L2 around the same age, i.e., 3 to 5 years of age.

I argued that the asymmetry found in the segmentation preferences of L1Basque and L1Spanish bilinguals might result from differences in proficiency between these two groups of bilinguals as a result of the diglossic situation characteristic of the Basque Country, in which Spanish is the prevailing language. Research on speech processing has revealed proficiency effects in the processing of the non-native language. As discussed in Chapter 3, studies on lexical access have shown that achieving a very high proficiency in a non-native language might entail differences in processing. Costa and Santesteban (2004) and Costa, Santesteban and Ivanova (2006) reported symmetrical language switching costs during lexical access in highly proficient bilinguals, as opposed to the asymmetrical language switching costs found in studies with lower proficient bilinguals (Meuter and Allport, 1999; Costa and Santesteban, 2004). Similarly, electrophysiological studies on syntactic processing of the non-native language have revealed that highly proficient bilinguals show the same ERPs signature as native speakers when processing certain syntactic violations, unlike low proficient bilinguals (Kotz, Holcomb and Osterhout, 2008; Rossi, Gugler, Friederici and Hahne, 2006). It is hence possible that L1Spanish bilinguals were not able to deploy the segmentation strategies of their two languages due to a lower proficiency in their L2, whereas L1Basque bilinguals, highly proficient in their two languages, were able to implement both.

Alternatively, I suggested another tentative explanation to the asymmetry obtained in the groups of L1Basque and L1Spanish bilinguals, i.e., that the segmentation strategy attributed to Basque, namely, frequent-final segmentation, might be the marked strategy, and can hence only be implemented if acquired in the native language. If, as proposed by Kayne (1994), all languages share a universal underlying word order in which heads precede their complements, the frequent-initial segmentation strategy might be the default procedure. L1Basque bilinguals would thus be able to deploy both the marked and unmarked strategies, whereas L1Spanish bilinguals would only have acquired the unmarked strategy characteristic of their L1, i.e., frequent-initial segmentation. Indeed, Zawiszeswki, Gutiérrez, Fernández and Laka (2011) reported different ERPs signatures between L1Basque and L1Spanish highly proficient bilinguals in the processing of violations of the Head-Directionality parameter in Basque, i.e., L1Spanish bilinguals did not attain native-like processing of the syntactic violations.

Data on the processing of violations of the Head-Directionality parameter in Spanish by L1Basque and L1Spanish bilinguals is yet unavailable. However, if, as suggested, the head-initial segmentation strategy is the unmarked strategy and can hence be acquired regardless of this strategy not being part of the native language's repertoire of processing procedures, it is plausible that no differences might emerge between L1Basque and L1Spanish bilinguals in the processing of violations of the Head-Directionality parameter in Spanish.

Finally, if native-like processing of the two frequency-based segmentation strategies is only achieved when the marked strategy is a property of the L1, this would entail that these segmentation strategies have a low degree of plasticity, i.e., that language-experience has a great effect in the processing of this information (Sanders, Neville and Woldorff, 2002), as found in the processing of certain syntactic structures and unlike in the processing of lexical and semantic information, in which native-like processing can be achieved by late though highly proficient learners (Sanders, Neville and Woldorff, 2002, Weber-Fox and Neville, 1996; Zawiszeswki, Gutiérrez, Fernández and Laka, 2011).

Experiment C (Chapter 5) aimed to examine the interaction of prosodic and statistical cues in the segmentation of an artificial language by Basque/Spanish bilinguals. As presented in Chapter 5, it has been proposed that changes in pitch are the acoustic correlate of head-final structures (Nespor et al, 2008). Additionally, Bion et al. (2011) and Nespor et al. (2008) proposed that pitch changes lead to prominent-initial grouping of the stimuli, following the Iambic/Trochaic Law (Hayes, 1995). In Experiment C, I created an artificial language similar to the one used in Experiment B but with the insertion of pitch changes in the stream, such that non-frequent syllables displayed a higher f0 (120 Hz) than frequent syllables (100 Hz). According to Nespor et al.'s (2008) hypothesis, such a stream should lead to a segmentation pattern in which infrequent syllables would be perceived as the initial syllables of a higher prosodic constituent, which is tantamount to a frequent-final segmentation pattern. The hypothesis to be tested in this experiment would be whether this is so, and whether there are differences between L1Basque and L1Spanish speakers in terms of degrees of preferences towards frequent-final segmentation.

Given that pitch changes are the proposed correlate to head-final structures, frequent-final segmentation of the language was hence predicted in L1Basque/L2Spanish bilinguals (i.e., native speakers of a head-final language). Frequent-final segmentation of the language in L1Spanish/L2Basque bilinguals would entail that these bilinguals had implemented the segmentation procedure of their L2.

The results of Experiment C showed that the insertion of this prosodic cue did not lead to different segmentation preferences in neither of the two groups of bilinguals, i.e., their segmentation preference did not differ from the preference obtained in a similar but prosodically flat artificial language (i.e., Experiment B). Additionally, no differences were obtained in the segmentation preferences between the two groups of bilinguals.

A greater frequent-final segmentation preference as compared to Experiment B in the group of L1Basque, but not L1Spanish bilinguals, would have signalled that L1Spanish bilinguals could not deploy the prosodic segmentation strategy of their L2. A greater frequent-final segmentation preference than in Experiment B in both groups of bilinguals would have suggested that the L1Spanish bilinguals could deploy the segmentation strategies of their two languages. However, the absence of a segmentation preference found in both groups of participants is uninformative as to whether bilingual speakers are able to implement the segmentation strategy of both their L1 and L2.

This apparent insensitivity to the prosodic cue in both groups of bilinguals prevents us from drawing any conclusions on the potential abilities of L1Spanish/L2Basque bilinguals to implement the segmentation strategy of the L2. The potential causes for this insensitivity to the prosodic cue will be discussed in the next section.

6.2 Statistical and prosodic cues within the hierarchy of segmental cues

One of the aims of the experiments in the present dissertation was to investigate the relative weight that adult speakers give to statistical and prosodic cues, as compared to other segmentation cues. Experiments A and B focused on the role of statistical cues in speech segmentation and the interplay of statistical and acoustic-phonetic cues, whereas Experiment C investigated the relative weight given by adult speakers to prosodic cues, as compared to statistical cues.

As described in Chapter 2, section 2.3, statistical and prosodic cues appear to belong to the lowest tiers in the hierarchy of segmentation cues proposed by Mattys, White and Melhorn (2005) for adult speakers. As shown in studies that pitted segmentation cues against each other, adult speakers appear to rely more heavily on lexical information than in any other segmentation cue (Mattys et al., 2005). Whenever lexical information is unavailable (e.g., in artificial language learning experiments) adult speakers weigh segmental cues (i.e., acoustic-phonetic and phonotactic cues) more heavily than prosodic and statistical cues (Finn and Hudson Kam, 2008; Mattys et al., 2005, McQueen, 1998). Adult speakers seem to make use of prosodic and statistical information as last resort cues when both lexical and segmental information are uninformative (Mattys et al., 2005). Last, prosodic cues appear to override statistical cues, which suggests that statistical cues occupy the lowest tier in the hierarchy of segmentation cues (Langus, Marchetto, Bion and Nespor, 2012; Shukla, Nespor and Mehler, 2007).

Experiment A explored adult speakers' sensitivity to statistical cues. In this experiment, a general frequent-final segmentation preference was obtained in all groups of participants (i.e., Basque/Spanish bilinguals and Spanish, English and French monolinguals). I argued that this bias might have been caused by the acoustic-phonetic properties of the German voice used to synthesize the stimuli. Specifically, I proposed that the VOT values of the German voiceless stops might be leading participants (none of which were speakers of German) to segment the stimuli at the onset of the voiceless stops, which occurred always at the onset of infrequent syllables, resulting hence in a frequent-final segmentation preference of the artificial language.

This proposal was supported by the general frequent-initial segmentation preference obtained in Experiment B. The artificial language in Experiment B had a structure similar to the artificial language in Experiment A, but was synthesized with a Spanish voice, i.e., a voice that contained phonemes common to the repertoires of the participants' native languages (i.e., Basque, Spanish), hence eliminating the acoustic-phonetic cues proposed as the cause for the general frequent-final segmentation obtained in Experiment A. The shift in segmentation preference displayed in all groups in Experiment B, as compared to Experiment A, suggests that the acoustic-phonetic cues present in the input, and not the statistical information, determined the participants' segmentation preference. This result tallies with the prevalence of acoustic-phonetic cues (i.e., coarticulation) over statistical cues obtained by Fernandes, Ventura and Kolinsky (2007) with adult speakers and it also provides supporting evidence to the hierarchical account of segmentation cues proposed by Mattys et al. (2005).

Although, as mentioned above, all participants in Experiment A preferred a general frequent-final segmentation of the language, differences in segmentation preferences were obtained between participants who were native speakers of OV languages (e.g., L1Basque bilinguals) and speakers of VO languages (e.g., Spanish, French and English monolinguals). This difference followed the predictions of the frequency-based bootstrapping hypothesis—which will be addressed in the next section— and suggests that, though mainly driven by the acoustic-phonetic properties of the input, the statistical information modulated nonetheless the participants' segmentation preferences, tallying with the results of numerous studies that have shown that segmentation is modulated by the segmentation cues that belong to the lower tiers of the hierarchy, available to adult speakers (Andruski, Blumstein and Burton, 1994; Dumay, Content and Frauenfelder, 1999; McQueen, Norris and Cutler, 1994; Norris, McQueen and Cutler, 1995; Sanders and Neville, 2000; Smith and Hawkins, 2000; Vroomen, van Zon and de Gelder, 1996).

In Experiment C, participants listened to an artificial language that consisted of alternating frequent and infrequent syllables similar to the language in Experiment B, and hence containing the same statistical information. In addition to this statistical cue, the artificial language in Experiment C contained a prosodic cue to the segmentation of the language: alternating pitch rises and falls. As mentioned in the previous section, Bion et al. (2011) and Nespor et al. (2008) claim that this prosodic cue leads to a prominent-initial segmentation of a string of sounds, i.e., a frequent-final segmentation of the stream in Experiment C. Both statistical and prosodic information should cue to a frequent-final segmentation of the language to speakers of OV languages. On the contrary, these two cues would provide conflicting information to speakers of VO languages: the statistical cue signalled a frequent-initial segmentation whereas prosodic information cued to a frequent-final segmentation of the language. A difference in the segmentation preferences of L1Basque and L1Spanish bilinguals would thus reveal that participants had integrated these two cues. However, no such difference obtained. Furthermore, as mentioned in the previous section, prosodic information did not significantly influence the segmentation preferences of neither group of participants.

These results suggest that participants were not relying on prosodic cues to segment the input. Instead, participants displayed a similar frequent-initial segmentation preference to that obtained in Experiment B. In Experiment B, I argued that the general frequent-initial bias found in all groups might result from potential acoustic-phonetic cues present in the input. Since the artificial languages in experiments B and C contained the same acoustic-phonetic information, the apparent insensitivity to prosodic cues observed in Experiment C would suggest that prosodic cues were outranked by segmental information, as predicted by Mattys et al. (2005). Nevertheless, it is impossible to determine which of these two situations was at play: (a) that integration of these prosodic and statistical cues, both ranked lower than segmental information in the hierarchy of cues, might not have sufficed to overrule potential segmental information present in the input, or (b) that participants were not able to integrate the two segmentation cues. This indeterminacy opens the ground for future investigations.

In Chapter 4, I proposed an alternative explanation for the general frequent-initial bias found in Experiment B, namely that this bias might reveal a default frequent-initial segmentation preference which would be universally preferred. In this scenario, the frequent-initial segmentation preference found in Experiment C would suggest that neither the prosodic information nor the convergent information provided by prosodic and statistical cues were powerful enough to overcome this default preference.

Though further research is required in order to determine whether the general frequent-initial segmentation preference found in experiments B and C results from the acoustic-phonetic information present of the input or reveals a universally preferred segmentation strategy, the absence of an effect caused by the prosodic cue introduced in Experiment C reveals that prosodic information plays a minor role in speech segmentation by adult speakers.

6.3 Availability in adult speech segmentation of frequency-based and prosodic cues proposed as bootstrapping mechanisms to the acquisition of syntax

The third aim of the experiments in the present dissertation was to examine whether adult speakers made use of statistical and prosodic cues in speech segmentation. Specifically, the statistical cue investigated was the frequency distribution and relative order of frequent and infrequent elements, mirroring the frequency distribution of functors and content words in natural languages. The prosodic cue examined was the acoustic realization and relative location of prominence within phonological phrases. It has been proposed that these two cues might play a crucial role in bootstrapping syntax acquisition, allowing infants to set the value of the Head-Directionality parameter prior to having lexical knowledge, due to the correlation of these two cues with the order of heads and complements in natural languages (Christophe, Guasti, Nespor, Dupoux and van Ooyen, 1997; Gervain, 2007; Nespor, Guasti and Christophe, 1996; Nespor, Shukla, van de Vijver, Avesani, Schraudolf and Donati, 2008).

As presented in Chapter 2, section 2.1.2, functors are highly frequent elements in natural languages, which tend to occur phrase-initially in head-initial languages and phrase-finally in head-final languages (Gervain, 2007). The frequency-based bootstrapping hypothesis (Gervain, 2007) proposes that infants track the frequency distribution of functors and their order of occurrence relative to content words, building a rudimentary representation of the order of heads and complements in the language under acquisition.

As presented in Chapter 3, Gervain (2007) conducted a series of artificial language learning experiments with adult and infant speakers of several head-initial (VO) and head-final (OV) languages and found that both adult and infants were able to track the frequency distribution of the elements in an ambiguous artificial language, and that their segmentation preferences appeared to correlate with the relative order of functors (frequent elements) and content words (infrequent elements) in their native languages. Thus, 8-month-old learners of a head-final language (i.e., Japanese) looked longer to frequent-final stimuli than to frequent-initial stimuli, whereas learners of a head-initial language (i.e., Italian) displayed the opposite pattern. Similarly, adult speakers of head-final languages (i.e., Basque, Japanese, Hungarian) choose a frequent-final segmentation of the language, whereas speakers of head-initial languages (i.e., French, Italian) chose a frequent-initial segmentation of the language.

Experiment A in the present dissertation aimed to further explore the availability of the frequency-based segmentation strategies in adult speakers of head-initial and head-final languages. To that end, two groups of native speakers of previously untested languages (English and Spanish, both head-initial VO languages), as well as native speakers of two languages already tested by Gervain (Basque, OV, and French, VO) were presented with the same ambiguous artificial language created and used by Gervain (2007)[36].

As predicted by the frequency-based bootstrapping hypothesis, speakers of the OV language (i.e., Basque) chose a frequent-final segmentation of the language more often than speakers of the VO languages (i.e., French, Spanish and English). This result suggests that adult speakers tracked the frequency cue contained in the input and that speakers of VO and OV languages had different representations of the relative order of frequent and infrequent elements which correlated with the relative order of functors and content words in their native languages. These different segmentation preferences replicated hence Gervain's (2007) results and provided supporting evidence to the proposal that the frequency distribution of elements in the input constitute a useful cue to speech segmentation and that frequent elements in natural languages (i.e., functors) act as anchoring points to syntactic structure.

The speakers of VO languages in Experiment A did however not display the expected frequent-initial segmentation preference reported in Gervain (2007), but a frequent-final segmentation preference, though smaller than the preference observed in speakers of OV languages, as a result of a bias caused by the acoustic-phonetic properties of the German voice used to synthesize the stimuli—as discussed in the previous section. Acoustic-phonetic cues overruled hence the frequency cues contained in the input.

As presented in Chapter 2, section 2.3, infants approach the hierarchy of segmental cues bottom-up, instead of the top-down access that characterizes adult speakers. At the onset of language acquisition, infants rely on statistical and prosodic information, prior to the acquisition of segmental cues and lexical knowledge (Johnson

[36] Identical in structure, lexicon, familiarization and test stimuli, but synthesized with a German voice instead of the Spanish voice used by Gervain (2007).

and Tyler, 2010; Jusczyk, Houston and Newsome, 1999). Therefore, the fact that the segmentation preferences of adult speakers were modulated (though not determined) by the frequency-based segmentation cues, in addition to the salient properties of functors in natural languages and the crucial role that statistical information appears to play at the onset of language acquisition, give plausibility to Gervain's (2007) proposal that the frequency-based cues might allow infants bootstrap a major syntactic parameter, i.e., the Head-Directionality parameter.

In Experiment B, native speakers of Basque (OV) and Spanish (VO) listened to a new artificial language, which had the same structure as the language used in Experiment A and Gervain (2007) but was synthesized with the same Spanish voice used in the experiments in Gervain (2007), instead of the German voice used in Experiment A. This change in the voice aimed to examine whether speakers of Spanish (VO) displayed the frequent-initial segmentation predicted by the frequency-based bootstrapping hypothesis, once the acoustic-phonetic information that caused the general frequent-final bias observed in Experiment A was eliminated.

As predicted, native speakers of Spanish (i.e., Spanish monolinguals and L1Spanish bilinguals) preferred a frequent-initial segmentation of the language. Surprisingly, L1Basque participants displayed a similar preference for frequent-initial segmentation, contrary to the frequent-final segmentation preference that is predicted by this hypothesis and that was found in Gervain (2007) and in Experiment A, with a similar population. As discussed in Chapter 4, this frequent-initial segmentation preference is striking, given that the artificial language in Experiment B had the same structure and was synthesized with the same voice as the artificial language used in Gervain (2007).

As mentioned in the previous section, I proposed two tentative explanations for the general frequent-initial segmentation preference obtained, i.e., the influence of acoustic-phonetic cues, and the possibility that frequent-initial segmentation is the universally preferred segmentation. The lexicon of the artificial language in Experiment B contained two phonemes that did occur in Gervain (2007) and Experiment A's lexicon, namely /l/ and /s/.[37] It seems unlikely that these two segments, which form part

[37] These phonemes were included in substitution of other phonemes present in the lexicon in Experiment A, in order to avoid potential phonotactic and allophonic effects (as described in the methodology of Experiment B in Chapter 4).

of the phonological repertoires of both Basque and Spanish, might have caused this bias. However, this possibility would have to be ruled out by means of a control experiment in which the lexicon would only contain phonemes that were also contained in Gervain's language (2007). On the other hand, the general frequent-initial segmentation found in Experiment B and C provides supporting evidence to the proposal that frequent-initial segmentation might be the default segmentation procedure. However, this proposal cannot account for the frequent-final segmentation preference obtained in Gervain (2007) with speakers of Japanese and Basque (OV).

Therefore, I cannot successfully account for the cause of the general frequent-initial segmentation preference obtained, or rather for the absence of a frequent-final segmentation in native Basque speakers. Further investigations that will aim to clarify this issue are pending.

Experiment C aimed to investigate a prosodic cue that several authors working on the phonological bootstrapping hypothesis have suggested that helps infants set the value of the Head-Directionality parameter. This cue is the location and realization of the main prominence within phonological phrases (Christophe, Guasti, Nespor, Dupoux and van Ooyen, 1997; Nespor, Guasti and Christophe, 1996; Nespor, Shukla, van de Vijver, Avesani, Schraudolf and Donati, 2008). Specifically, experiment C aimed to test whether Basque/Spanish bilinguals were sensitive to an acoustic cue associated to the word order characteristic of a head-final language like Basque.

As presented in Chapter 2, section 2.2, adults and infants are sensitive to the prosodic information contained in the input. Infants' sensitivity to prosodic cues is found at the earliest stages of development, and these cues might hence be crucial to the bootstrapping of the lexicon and syntax. The location of the main prominence within phonological phrases is correlated with the relative order of heads and complements in natural languages (Nespor and Vogel, 1986), and it has been proposed that this prominence has different acoustic realizations in head-initial and head-final structures (Nespor et al., 2008), following the Iambic/Trochaic Law (Hayes, 1995). In head-initial languages, prominence falls on the rightmost word of the phonological phrase and is realized by increased duration, whereas in head-final languages main prominence falls on the leftmost word and is realized by means of higher pitch (Nespor and Vogel, 1986; Nespor et al., 2008). Bion, Benavides-Varela and Nespor (2011) showed that adult speakers grouped stimuli that contained changes in pitch into prominent-initial groupings, and stimuli that contained changes in duration into prominent-final groupings.

In order to test the bilingual speakers' sensitivity to this cue, Basque/Spanish bilingual participants were presented with a similar artificial language to the one used in Experiment B, with the added manipulation of increased pitch in each infrequent syllable, hence signalling frequent-final segmentation of the language. However, neither group of participants chose this expected frequent-final segmentation. Therefore, the results obtained in Experiment C did not replicate the prominent-initial grouping obtained in Bion et al. (2011) in stimuli that contained changes in pitch. Note however that, unlike in Bion et al. (2011), the artificial language in Experiment C additionally contained a statistical cue.

The experiments in the present dissertation show that the frequency-based segmentation strategies, which Gervain (2007) proposes might help infants bootstrap syntactic acquisition, are also available to adult speakers, though easily overridden by other cues present in the input. On the other hand, the results obtained in Experiment C show that adult speakers do not make use of the prosodic cue proposed by Nespor et al. (2008) to assist infants in bootstrapping syntax acquisition. The influence of these two cues seems hence to be at best limited in adult speech segmentation. However, as described above, statistical and prosodic cues appear to play a crucial role during the earliest stages of language acquisition. These two cues are the first cues picked up by infants, who start integrating statistical and prosodic information at around 8 to 9 months of age (Johnson and Tyler, 2010; Jusczyk, Houston and Newsome, 1999; Morgan, 1994; Morgan and Saffran, 1995).

Indeed, it has been proposed that bootstrapping syntax acquisition might result from the integration of these two sources of information, i.e., statistical and prosodic cues (Gervain, 2007; Shukla and Nespor, 2010). Christophe, Millotte, Bernal and Lidz (2008) propose that phrasal prosody and functional elements might allow infants to conduct a first pass analysis of the input, bootstrapping the acquisition of syntactic structure and syntactic categories. According to Christophe et al. (2008), prosodic information would assist infants in the location of phrase boundaries, while functors would help infer the syntactic category of the adjacent words within the syntactic constituents.

As presented in Chapter 2, 6- to 9-month-old infants are sensitive to the prosodic markers of phonological phrases, i.e., infants prefer to listen to prosodically well-formed phrases than to ill-formed phrases (Jusczyk et al., 1992; Soderstrom et al., 2003). Also, word segmentation is limited to the domains of phonological phrases, i.e., infants do not treat sequences that straddle phonological phrase boundaries as word candidates (Gout, Christophe and Morgan, 2004). Similarly, sensitivity to functional elements appears very early in development, due to their salient statistical, phonological and distributional properties. By 8 months of age, not only do infants segment functors from fluent speech and have detailed representation of at least the highest frequency functors, but infants make use of these functors to segment adjacent content words (Höhle and Weissenborn, 2003; Shi, Cutler, Werker and Cruishank, 2006; Shi, Marquis and Gauthier, 2006). By 24 months of age infants have already gathered some knowledge of the co-occurrence patterns between functional and content word categories (Gerken and McIntosh, 1993). Functors, which tend to occur at the edges of syntactic phrases, have been proposed as markers that help learn the syntactic rules of the languages (Green, 1979; Valian and Coulson, 1988).

It is hence plausible that the joint information provided by functors and phrasal prosody might allow infants to bootstrap syntactic acquisition. Phrasal prosody would allow infants to divide the input into chunks characterized by the presence of functors at the edges of these smaller chunks. The distributional properties of functors might in turn enable infants to discover the co-occurrence patterns of specific functor and content word categories.

Indeed, it has been proposed by Endress and Mehler (2009) that edges are perceptual primitives that help listeners encode information. Research on artificial language learning has revealed that adult participants successfully learn a non-adjacent dependency when the dependency occurs at edge position, but not when the dependency occurs sequence-internally (Endress and Mehler, 2009). Similarly, adult speakers learn new phonotactic constraints when the constraints occur at the edges, (e.g., C_1VccVC_2), but do not learn these new constraints when the critical segments occur word-internally (e.g., cVC_1C_2Vc), unless provided with an additional cue (e.g., when the set of consonants allowed in *C1* and *C2* belong to different natural classes, Endress and Mehler, 2010). Also, adult speakers generalize repetition-based structures when the repetitions occur at the edges of the sequences (e.g., ABCDE*<u>FF</u>*), but not sequence-internally (e.g., ABC*<u>DD</u>*EF) (Endress, Scholl and Mehler, 2005). Furthermore, as described in Chapter 2, in natural languages main prominence falls at the edges of the phonological phrases. Also, as mentioned by Endress and Mehler (2009), affixes in natural languages do generally occur at the edges of words (i.e., suffixes, prefixes) rather than word internally. Last, studies on the input strategies that mothers use to teach new words to infants have revealed that mothers tend to place the new information at the final edge of the utterances (Woodward and Aslin, 1990, 1991, 1992, as cited in Aslin, Woodward, LaMendola and Bever, 1996).

6.4 Conclusions

→ Bilingual speakers are able to deploy frequency-based segmentation strategies that characterize their two languages, though acquisition of the L2's segmentation strategy appears to be constrained. The causes of this constraint are yet to be determined.

→ Statistical and prosodic cues seem to play a minor role in speech segmentation for adult speakers. That is, statistical and prosodic cues are outranked by other segmentation cues such as acoustic-phonetic cues, supporting therefore a hierarchical account of segmentation cues in which statistical and prosodic information are the least weighed segmentation cues by adult speakers.

→ Frequent-initial segmentation might be the universally preferred segmentation strategy. Further research that aims to confirm this hypothesis is pending.

→ Frequency-based segmentation strategies are available cues to adult speakers in speech segmentation. That is, adult speakers track the frequency distribution of elements, and speakers' segmentation preferences are modulated by the relative order of frequent (functors) and infrequent elements (content words) in their native languages.

Appendix 1.
Background questionnaire

The following questionnaire is an English translation of the original questionnaire in Spanish and Basque. The originals were displayed in Excel sheets and included multi-option cells, which are here presented in tables.

PLEASE, FILL OUT THE FOLLOWING QUESTIONNAIRE.

Experimenter:

Code of the experiment:

Date:

Time:

Name:

Family name:

Gender:

Birth date:

e-mail address:

Telephone:

Place of residence (town and province):

Place of birth (town and province):

Have you ever lived in another town and/or province?

If so, in which town and province?

From which date to which date?

Background questionnaire

• *As an infant, in which language were you spoken to by...*
(multi-option, see Table 18)

...you mother?
...your father?
...your brother(s) and sister(s)?
...your grandparents?

Table 18. Options given to the participants

Only Spanish.
Mainly Spanish, albeit sometimes Basque.
Basque and Spanish with equal frequency.
Mainly Basque, albeit sometimes Spanish.
Only Basque.

• *At what age did you start...*
...speaking Spanish?
...speaking Basque?

• *If you have a language proficiency certificate on Basque, please specify which one:*

• *Which language and with which frequency do/did you use...*
(multi-option, see Table 19)

...as an infant, prior to going to school?
...as an infant, during primary school, at school?
...as an infant, during primary school, at home?
...as an infant, during primary school, elsewhere?

...as a teenager, during secondary school, at high school?
...as a teenager, during secondary school, at home?
...as a teenager, during secondary school, elsewhere?

...at the moment, in adulthood, at the university/work?

...at the moment, in adulthood, at home?

...at the moment, in adulthood, elsewhere?

Table 19. Options given to the participants

Only Spanish.
Almost always Spanish, rarely Basque.
Mainly Spanish, albeit using Basque at least 25% of the time.
Basque and Spanish with equal frequency.
Mainly Basque, albeit using Spanish at least 25% of the time.
Almost always Basque, rarely Spanish.
Only Basque.

• *In general, which language do you feel more comfortable in?*
(multi-option, see Table 20)

Table 20. Options given to the participants

Spanish.
Basque.
I feel comfortable in both languages.

• *With how much frequency do you read in Spanish/Basque?*
(multi-option, see Table 21)

Table 21. Options given to the participants

Never.
Little.
Neither too much nor too little.
Quite a lot.
Very much.

• *How do you get by in the following languages? Do you have any language proficiency certificates?*
(multi-option, see Table 22)

	speak	*understand*	*read*	*write*	*Certificate*
Basque					
Spanish					
English					
French					
Any other language? Which one?					

Table 22. Options given to the participants

Not at all
Very bad
Bad
So-so
Quite
Well
Very well
Perfectly

• *Choose your preferred hand when carrying out the following actions:*
(multi-option, see Table 23)

write
draw
throw an object
cut with scissors
brush your teeth
cut with a knife
use the spoon
sweep (hand placed above)
play tennis (service)
open a tin (hand placed above)

Table 23. Options given to the participants

Left hand
Right hand
Both hands

Appendix 2. Test items and counterbalanced test pairs in Experiment B

Table 24. Test items in Experiment B

Functor category	**Frequent-initial items**	**Frequent-final items**
fi	**fi**kanufupepo	tinunapeku**fi**
	filunusapeki	kanulipepo**fi**
	finenulipeta	fenukopema**fi**
	fisonumopepi	sonufupele**fi**
	fitinutepeku	lunutepepi**fi**
	fitonupupefo	panulopetu**fi**
pe	**pe**fofimununa	kufinenumo**pe**
	pelefipanufu	fofikanusa**pe**
	pekififenuko	mafitonuna**pe**
	petufinenupu	tafimunuli**pe**
	pemafikanulo	lefifenupu**pe**
	pepofitinusa	kifipanuko**pe**
nu	**nu**mopetafife	fupetafine**nu**
	nutepemafilu	lopepifito**nu**
	nulopepifimu	sapekifimu**nu**
	nukopetufipa	mopepofilu**nu**
	nunapelefiso	pupefofiso**nu**
	nulipekufito	tepetufiti**nu**

Table 25. Counterbalanced test pairs in Experiment B

	Pairs	
Combination	**1st member / 2nd member**	**1st member / 2nd member**
nu_ _nu	**nu**mopetafife / pupefofiso**nu**	**nu**kopetufipa / mopepofilu**nu**
nu_ _pe	**nu**tepemafilu / lefifenupu**pe**	**nu**lipekufito / kufinenumo**pe**
nu_ _fi	**nu**lopepifimu / fenukopema**fi**	**nu**napelefiso / panulopetu**fi**
pe_ _pe	**pe**fofimununa / mafitonuna**pe**	**pe**tufinenupu / tafimunuli**pe**
pe_ _nu	**pe**lefipanufu / lopepifito**nu**	**pe**mafikanulo / sapekifimu**nu**
pe_ _fi	**pe**kififenuko / sonufupele**fi**	**pe**pofitinusa / lunutepepi**fi**
fi_ _fi	**fi**kanufupepo / tinunapeku**fi**	**fi**sonumopepi / kanulipepo**fi**
fi_ _pe	**fi**nenulipeta / fofikanusa**pe**	**fi**tonupupefo / kifipanuko**pe**
fi_ _nu	**fi**lunusapeki / tepetufiti**nu**	**fi**tinutepeku / fupetafine**nu**
_nu nu_	fupetafine**nu / nu**tepemafilu	sapekifimu**nu / nu**napelefiso
_nu pe_	mopepofilu**nu / pe**tufinenupu	lopepifito**nu / pe**pofitinusa
_nu fi_	pupefofiso**nu / fi**nenulipeta	tepetufiti**nu / fi**sonumopepi
_pe pe_	kufinenumo**pe / pe**kififenuko	lefifenupu**pe / pe**mafikanulo
_pe nu_	fofikanusa**pe / nu**mopetafife	kifipanuko**pe / nu**lopepifimu
_pe fi_	mafitonuna**pe / fi**kanufupepo	tafimunuli**pe / fi**tinutepeku
_fi fi_	fenukopema**fi / fi**lunusapeki	tinunapeku**fi / fi**tonupupefo
_fi pe_	panulopetu**fi / pe**fofimununa	kanulipepo**fi / pe**lefipanufu
_fi nu_	sonufupele**fi / nu**kopetufipa	lunutepepi**fi / nu**lipekufito

Appendix 3. Statistical analysis of the frequency of CV syllables in Experiments A and B

The syllabic frequencies in Basque and Spanish were taken from the online application SYLLABARIUM (Duñabeitia, Cholin, Corral, Perea and Carreiras, 2010). The analyses were conducted of both token and type frequencies. All frequencies were converted into log_{10} values for analysis.

(i.) I compared the syllabic frequency at word-initial position of the CV syllables in experiments A vs. B, both in Basque and Spanish, by means of independent-sample t-tests.

• Comparison of the syllabic frequency of the CV syllables at initial position in Basque, in Experiment A vs. B:

→ The analysis of the token frequencies in initial position between Experiment A (M = 2.66; SD = 1.09) and Experiment B (M = 2.75; SD = .52) revealed no differences (t(58) =-.406, p = .686).

→ The analysis of the type frequencies in initial position between Experiment A (M = 1.73; SD = .66) and Experiment B (M = 1.86; SD = .33) revealed no differences (U= 437, Z=-1.92, p = .848).[38]

[38] The Kolmogórov-Smirnoff test of normality revealed that the dependent variable did not have a normal distribution (p = .033). Hence, I conducted a Mann-Whitney test instead of a t-test.

• Comparison of the syllabic frequency of the CV syllables at initial position in Spanish, in Experiment A vs. B:

→ The analysis of the token frequencies in initial position between Experiment A (M = 3.06; SD = .49) and Experiment B (M = 3.10; SD = .47) revealed no differences (t(58) =-.268, p = .789).

→ The analysis of the type frequencies in initial position between Experiment A(M = 2.17; SD = .40) and Experiment B (M = 2.17; SD = .32) revealed no differences (t(58) =-.025, p = .980).

• Comparison of the syllabic frequency of the subset of new CV syllables in Experiments B and the respective subset of original syllables in Experiment A, at initial position in Basque, in Experiment A vs. B:

→ The analysis of the token frequencies in initial position between Experiment A (M = 2.53; SD = 1.58) and Experiment B (M = 2.74; SD = .56) revealed no differences (t(24) =-.444, p = .661).

→ The analysis of the type frequencies in initial position between Experiment A (M = 1.56; SD = .92) and Experiment B (M = 1.86; SD = .34) revealed no differences (t(24) =-1.083, p = .290).

• Comparison of the syllabic frequency of the subset of new CV syllables in Experiments B and the respective subset of original syllables in Experiment A, at initial position in Spanish, in Experiment A vs. B:

→ The analysis of the token frequencies in initial position between Experiment A (M = 2.91; SD = .43) and Experiment B (M = 2.99; SD = .40) revealed no differences (t(24) =-.471, p = .642).

→ The analysis of the type frequencies in initial position between Experiment A (M = 2.09; SD = .48) and Experiment B (M = 2.09; SD = .31) revealed no differences (t(24) =-.034, p = .973).

(ii.) Additionally, I compared the syllabic frequency at all positions of the CV syllables in experiments A vs. B, both in Basque and Spanish, by means of t-tests.

• Comparison of the syllabic frequency of the CV syllables at all positions in Basque, in Experiment A vs. B:

→ The analysis of the token frequencies in all positions between Experiment A (M = 3.69; SD = .56) and Experiment B (M = 3.55; SD = .64) revealed no differences (t(58) = .864, p = .391).

→ The analysis of the type frequencies in all positions between Experiment A (M = 2.60; SD = .39) and Experiment B (M = 2.59; SD = .46) revealed no differences (t(58) = .079, p = .937).

• Comparison of the syllabic frequency of the CV syllables at all positions in Spanish, in Experiment A vs. B:

→ The analysis of the token frequencies in all positions between Experiment A (M = 3.69; SD = .56) and Experiment B (M = 3.65; SD = .47) revealed no differences (t(58) =-.392, p = .696).

→ The analysis of the type frequencies in all positions between Experiment A (M = 2.83; SD = .43) and Experiment B (M = 2.85; SD = .40) revealed no differences (t(58) =-.213, p = .832).

• Comparison of the syllabic frequency of the subset of new CV syllables in Experiments B and the respective subset of original syllables in Experiment A, in all positions in Basque, in Experiment A vs. B:

→ The analysis of the token frequencies in all positions between Experiment A (M = 3.97; SD = .34) and Experiment B (M = 3.66; SD = .69) revealed no differences (t(24) = 1.451, p = .160).

→ The analysis of the type frequencies in all positions between Experiment A (M = 2.69; SD = .28) and Experiment B (M = 2.67; SD = .48) revealed no differences (t(24) = .131, p = .897).

• Comparison of the syllabic frequency of the subset of new CV syllables in Experiments B and the respective subset of original syllables in Experiment A, in all positions in Spanish, in Experiment A vs. B:

→ The analysis of the token frequencies in all positions between Experiment A (M = 3.52; SD = .59) and Experiment B (M = 3.64; SD = .43) revealed no differences ($t(24) = -.588$, $p = .562$).

→ The analysis of the type frequencies in all positions between Experiment A (M = 2.82; SD = .46) and Experiment B (M = 2.87; SD = .39) revealed no differences ($t(24) = -.313$, $p = .757$).

Appendix 4. Statistical analysis of the frequency of use of Basque/Spanish reported by L1Basque and L1Spanish bilingual participants in Experiments A and B

1. (i.) I compared the frequency of use of Spanish/Basque outside from school/home reported by the group of L1Basque which received the instructions in Basque (i.e., context Basque), and the group of L1Basque which received the instructions in Spanish (i.e., context Spanish) in **Experiment A**, by means of contingency tables.

→ The comparison of the frequency of use of Spanish/Basque during infancy in L1Basque context Basque and L1Basque context Spanish bilinguals in Experiment A revealed no differences: $\chi^2(3, N = 48) = .52, p = .915$.

→ The comparison of the frequency of use of Spanish/Basque during adolescence in L1Basque context Basque and L1Basque context Spanish bilinguals in Experiment A revealed no differences: $\chi^2(4, N = 48) = 1.82, p = .770$.

→ The comparison of the frequency of use of Spanish/Basque in adulthood in L1Basque context Basque and L1Basque context Spanish bilinguals in Experiment A revealed no differences: $\chi^2(5, N = 48) = 4.51, p = .479$.

(ii.) I compared the frequency of use of Spanish/Basque outside from school/home reported by the group of L1Spanish context Basque, and the group of L1Spanish context Spanish in Experiment A, by means of contingency tables.

→ The comparison of the frequency of use of Spanish/Basque during infancy in L1Spanish context Basque and L1Spanish context Spanish bilinguals in Experiment A revealed no differences: $\chi^2(4, N = 63) = 1.07, p = .899$.

→ The comparison of the frequency of use of Spanish/Basque during adolescence in L1Spanish context Basque and L1Spanish context Spanish bilinguals in Experiment A revealed no differences: $\chi^2(5, N = 63) = 2.51, p = .774$.

→ The comparison of the frequency of use of Spanish/Basque in adulthood in L1Spanish context Basque and L1Spanish context Spanish bilinguals in Experiment A revealed no differences: $\chi^2(5, N = 63) = 3.58, p = .612$.

(iii.) I compared the reported frequency of use of Spanish/Basque outside from school/home of L1Basque and L1Spanish bilinguals in Experiment A, by means of contingency tables.

→ The comparison of the frequency of use of Spanish/Basque during infancy in L1Basque and L1Spanish bilinguals in Experiment A revealed significant differences: $\chi^2(6, N = 111) = 96.11, p < .001$.

→ The comparison of the frequency of use of Spanish/Basque during adolescence in L1Basque and L1Spanish bilinguals in Experiment A revealed significant differences: χ^2 $(6, N = 111) = 73.01, p < .001$.

→ The comparison of the frequency of use of Spanish/Basque in adulthood in L1Basque and L1Spanish bilinguals in Experiment A revealed significant differences: $\chi^2(6, N = 111) = 52.52, p < .001$.

2. (i.) I compared the frequency of use of Spanish/Basque outside from school/home reported by the group of L1Basque context Basque and the group of L1Basque context Spanish in **Experiment B**, by means of contingency tables.

→ The comparison of the frequency of use of Spanish/Basque during infancy in L1Basque context Basque and L1Basque context Spanish bilinguals in Experiment B revealed no differences: $\chi^2(4, N = 50) = 5.30, p = .258$.

→ The comparison of the frequency of use of Spanish/Basque during adolescence in L1Basque context Basque and L1Basque context Spanish bilinguals in Experiment B revealed no differences: $\chi^2(4, N = 50) = 6.03, p = .197$.

→ The comparison of the frequency of use of Spanish/Basque in adulthood in L1Basque context Basque and L1Basque context Spanish bilinguals in Experiment B revealed no differences: $\chi^2(4, N = 50) = 6.40, p = .171$.

(ii.) I compared the frequency of use of Spanish/Basque outside from school/home reported by the group of L1Spanish context Basque, and the group of L1Spanish context Spanish in Experiment B, by means of contingency tables.

→ The comparison of the frequency of use of Spanish/Basque during infancy in L1Spanish context Basque and L1Spanish context Spanish bilinguals in Experiment B revealed no differences: $\chi^2(6, N = 42) = 6.53, p = .367$.

→ The comparison of the frequency of use of Spanish/Basque during adolescence in L1Spanish context Basque and L1Spanish context Spanish bilinguals in Experiment B revealed no differences: $\chi^2(5, N = 42) = 7.18, p = .208$.

→ The comparison of the frequency of use of Spanish/Basque in adulthood in L1Spanish context Basque and L1Spanish context Spanish bilinguals in Experiment B revealed no differences: $\chi^2(5, N = 42) = 6.17, p = .290$.

(iii.) I compared the reported frequency of use of Spanish/Basque outside from school/home of L1Basque and L1Spanish bilinguals in Experiment B, by means of contingency tables.

→ The comparison of the frequency of use of Spanish/Basque during infancy in L1Basque and L1Spanish bilinguals in Experiment B revealed significant differences: $\chi^2(6, N = 92) = 65.97, p < .001$.

→ The comparison of the frequency of use of Spanish/Basque during adolescence in L1Basque and L1Spanish bilinguals in Experiment B revealed significant differences: $\chi^2(6, N = 92) = 64.64, p < .001$.

→ The comparison of the frequency of use of Spanish/Basque in adulthood in L1Basque and L1Spanish bilinguals in Experiment B revealed significant differences: $\chi^2(6, N = 92) = 61.75, p < .001$.

3. (i.) I compared the frequency of use of Spanish/Basque outside from school/home reported by L1Basque bilinguals in **experiments A and B**, by means of contingency tables.

→ The comparison of the frequency of use of Spanish/Basque during infancy in L1Basque bilinguals in Experiment A vs. Experiment B revealed no differences: $\chi^2(4, N = 98) = 2.98, p = .561$.

→ The comparison of the frequency of use of Spanish/Basque during adolescence in L1Basque bilinguals in Experiment A vs. Experiment B revealed no differences: $\chi^2(4, N = 98) = 6.77, p = .148$.

Statistical analysis of the bilinguals' frequency of use of Basque/Spanish

→ The comparison of the frequency of use of Spanish/Basque in adulthood in L1Basque bilinguals in Experiment A vs. Experiment B revealed no differences: $\chi^2(6, N = 98) = 4.94$, $p = .552$.

(ii.) I compared the frequency of use of Spanish/Basque outside from school/home reported by L1Spanish bilinguals in **experiments A and B**, by means of contingency tables.

→ The comparison of the frequency of use of Spanish/Basque during infancy in L1Spanish bilinguals in Experiment A vs. Experiment B revealed no differences: $\chi^2(6, N = 105) = 6.34$, $p = .386$.

→ The comparison of the frequency of use of Spanish/Basque during adolescence in L1Spanish bilinguals in Experiment A vs. Experiment B revealed no differences: $\chi^2(6, N = 105) = 4.27$, $p = .641$.

→ The comparison of the frequency of use of Spanish/Basque in adulthood in L1Spanish bilinguals in Experiment A vs. Experiment B revealed no differences: $\chi^2(5, N = 105) = 3.02$, $p = .697$.

Appendix 5.

Traducción del capítulo 6: Discusión general y conclusiones

En los capítulos 3, 4 y 5 de la presente tesis doctoral se describieron y discutieron tres experimentos sobre aprendizaje de lenguas artificiales (experimentos A, B and C), los cuales investigaban las habilidades de hablantes adultos monolingües y bilingües para emplear pistas estadísticas y prosódicas en la segmentación del input.

El objetivo principal de estos experimentos era examinar si los hablantes bilingües eran capaces de implementar las estrategias de segmentación de sus dos lenguas, o si, en cambio, empleaban únicamente las estrategias características de su L1 en la segmentación del habla. Además, los experimentos de la presente tesis tenían como objetivo explorar el peso relativo que los hablantes adultos confieren a las pistas estadísticas y prosódicas en el marco de la concepción jerárquica de las pistas de segmentación propuesto por Mattys, White y Melhorn (2005). Por último, estos experimentos buscaban observar si los hablantes adultos emplean un tipo de información estadística y un tipo de información prosódica, las cuales han sido propuestas como potencialmente cruciales para que los niños emprendan la adquisición de la sintaxis, a saber, la distribución de la frecuencia y el orden relativo de los elementos frecuentes e infrecuentes, reflejando así los elementos funcionales y de contenido en las lenguas naturales, y la realización acústica y localización relativa de la prominencia en los sintagmas fonológicos. Se ha propuesto que estas dos pistas se correlacionan con el orden de cabezas y complementos en las lenguas naturales, y podrían por tanto permitir a los niños construir una representación rudimentaria del orden de palabras, asignando el valor del principal parámetro configuracional, el parámetro de direccionalidad de la cabeza sintáctica. Estas tres cuestiones se abordan a continuación separadamente.

6.1 La implementación de procedimientos de segmentación en bilingües

Como se mencionó antes, el objetivo principal de esta tesis doctoral era explorar las habilidades de los hablantes bilingües para segmentar el habla fluida. Los experimentos A y B (capítulos 3 y 4) se centraban en una pista estadística para la segmentación del input, la frecuencia. Como se presentó en el capítulo 2, sección 2.1, tanto niños como adultos son extremadamente sensibles a la información estadística contenida en el input, información que emplean tanto para localizar las fronteras de palabras y sintagmas como para extraer regularidades (Gómez y Maye, 2005; Peña, Bonatti, Nespor y Mehler, 2002; Saffran, Aslin y Newport, 1996; Saffran, Newport y Aslin, 1996).

En los experimentos A y B presenté a hablantes bilingües de euskera/castellano con una lengua artificial ambigua y prosódicamente plana, caracterizada por una alternancia estricta de sílabas frecuentes e infrecuentes. Según las predicciones de la *frequency-based bootstrapping hypothesis* (Gervain, 2007)—a la cual me referiré más adelante—, se esperaba que hablantes de lenguas OV, de cabeza final, mostraran una mayor tendencia hacia una segmentación de la lengua ambigua en la cual los elementos frecuentes aparecen en posición final, en comparación con hablantes de lenguas VO, de cabeza inicial.

Los resultados de los experimentos A (capítulo 3) y B (capítulo 4) revelaron que las preferencias de segmentación de los hablantes bilingües L1euskera/L2castellano, estaban moduladas por la lengua de contexto (la lengua en la cual la investigadora se dirigió a los participantes y la lengua en la que recibieron las instrucciones) en la dirección predicha por la *frequency-based bootstrapping hypothesis*. Así, los hablantes bilingües L1euskera que recibieron las instrucciones en euskera eligieron una segmentación en la cual los elementos frecuentes aparecen en posición final con más frecuencia que los bilingües L1euskera a los cuales se habló en castellano. En cambio, la lengua de contexto no ejerció esta influencia en las preferencias de segmentación de la lengua artificial de los bilingües L1castellano/L2euskera.

Estos resultados sugieren que los bilingües L1euskera, a diferencia de los bilingües L1castellano, fueron capaces de emplear las estrategias de segmentación tanto de su L1 (euskera) como de su L2 (castellano). La edad de adquisición de la L2 es uno de los factores que se sabe influyen en el procesamiento de la L2 (Weber-Fox y Neville, 1996). Este factor no puede explicar la asimetría encontrada entre las habilidades de segmentación de los bilingües L1euskera y L1castellano, dado que ambos grupos de bilingües adquirieron su L2 respectiva alrededor de la misma edad, sobre los 3 a 5 años.

Argumenté que la asimetría encontrada en las preferencias de segmentación de los bilingües L1euskera y L1castellano podría resultar de diferencias en la competencia entre estos dos grupos de bilingües, como resultado de la situación de diglosia característica del País Vasco, región en la cual el castellano es la lengua preponderante. Investigación sobre el procesamiento del habla ha revelado efectos de competencia en el procesamiento de la lengua no nativa. Como se discutió en el capítulo 3, estudios sobre acceso léxico han mostrado que alcanzar una competencia muy alta en una lengua no nativa podría conllevar diferencias en el procesamiento. Costa y Santesteban (2004) y Costa, Santesteban e Ivanova (2006) encontraron costes simétricos en la alternancia de lenguas durante el acceso léxico en bilingües muy competentes, en oposición a los costes asimétricos de alternancia de lengua encontrados en estudios con bilingües menos competentes (Costa y Santesteban, 2004; Meuter y Allport, 1999). Igualmente, estudios electrofisiológicos sobre el procesamiento sintáctico de la lengua no nativa han revelado que bilingües altamente competentes muestran la misma firma electrofisiológica (ERPs) que hablantes nativos durante el procesamiento de ciertas violaciones sintácticas, a diferencia de bilingües poco competentes (Kotz, Holcomb y Osterhout, 2008; Rossi, Gugler, Friederici y Hahne, 2006). Por tanto, es posible que los bilingües L1castellano no hayan sido capaces de utilizar las estrategias de segmentación de sus dos lenguas debido a una menor competencia en su L2, mientras que los bilingües L1euskera, altamente competentes en sus dos lenguas, hayan sido capaces de implementar ambas estrategias.

Como alternativa, sugerí una segunda explicación tentativa a la asimetría obtenida entre los grupos de bilingües L1euskera y L1castellano: que la estrategia de segmentación atribuida al euskera, esto es, la segmentación en la cual los elementos frecuentes ocupan una posición final, podría ser la estrategia marcada, y podría por tanto implementarse únicamente si esta estrategia ha sido adquirida como parte de la lengua nativa. Si, tal y como propone Kayne (1994), todas las lenguas comparten un orden de palabras subyacente universal en el cual las cabezas preceden a sus complementos, la estrategia de segmentación de en la cual los elementos frecuentes ocupan una posición inicial podría ser el procedimiento por defecto. Los hablantes bilingües L1euskera podrían en consecuencia ser capaces de utilizar tanto la estrategia marcada como la no marcada, mientras que los hablantes L1castellano habrían adquirido únicamente la estrategia no marcada característica de su L1, esto es, la segmentación la cual los elementos frecuentes ocupan una posición inicial. De hecho, Zawiszeswki, Gutiérrez, Fernández y Laka (2011) encontraron firmas electrofisiológicas distintas en el procesamiento de violaciones del parámetro de direccionalidad de la cabeza sintáctica en euskera entre bilingües altamente competentes con L1euskera y L1castellano. Los bilingües L1castellano no alcanzaron un procesamiento de las violaciones sintácticas similar al encontrado en los hablantes nativos.

Por el momento no hay datos disponibles sobre el procesamiento de violaciones del parámetro de direccionalidad de la cabeza sintáctica en bilingües L1euskera y L1castellano. No obstante, si, como sugerí, la estrategia de segmentación de cabeza inicial es la estrategia no marcada y puede por tanto ser adquirida independientemente de el hecho de que esta estrategia no sea parte del repertorio de procedimientos de procesamiento de la lengua nativa, es plausible que no emerjan diferencias entre los bilingües con L1euskera y L1castellano en el procesamiento de violaciones del parámetro de direccionalidad de la cabeza sintáctica en castellano.

Finalmente, si un procesamiento similar al de un nativo de las dos estrategias de segmentación basadas en la frecuencia se alcanza únicamente cuando la estrategia marcada es una propiedad de la L1, esto implicaría que estas estrategias de segmentación tienen un grado bajo de plasticidad, esto es, que la experiencia con la lengua tiene un gran efecto en el procesamiento de esta información (Sanders, Neville y Woldorff, 2002), tal y como se ha encontrado en el procesamiento de ciertas estructuras sintácticas y a diferencia del procesamiento de la información léxica y semántica, en el cual puede alcanzarse un procesamiento similar al de un nativo incluso por aprendientes tardíos aunque altamente competentes (Sanders, Neville y Woldorff, 2002, Weber-Fox y Neville, 1996; Zawiszeswki, Gutiérrez, Fernández y Laka, 2011).

El objetivo del experimento C (capítulo 5) era examinar la interacción de pistas prosódicas y estadísticas en la segmentación de una lengua artificial por parte de bilingües de euskera/castellano. Como se presentó en el capítulo 5, se ha propuesto que los cambios en la altura tonal son el correlato acústico de las estructuras de cabeza final (Nespor et al. 2008). Además, Bion et al. (2011) y Nespor et al. (2008) propusieron que cambios en la altura tonal llevaban a una agrupación de los estímulos en la que la prominencia ocupaba una posición inicial, siguiendo la ley Yámbico/Trocaico (Hayes, 1995). En el experimento C, creé una lengua artificial similar a la empleada en el experimento B, aunque con la inserción en la cadena de cambios en la altura tonal, de forma que las sílabas no frecuentes estaban caracterizadas por tener una altura tonal mayor (120 Hz) que las sílabas frecuentes (100 Hz). Según la hipótesis de Nespor et al. (2008), una cadena de estas características debería inducir un patrón de segmentación en el cual las sílabas infrecuentes serían percibidas como las sílabas iniciales de un constituyente prosódico superior, lo que equivaldría a un patrón de segmentación en el cual los elementos frecuentes ocupan una posición final. En este experimento se examinó la validez de esta hipótesis, así como si surgían diferencias entre los hablantes L1euskera y L1castellano en términos de grados de preferencia hacia una segmentación en la cual los elementos frecuentes ocupan una posición final.

Dado que se ha propuesto que los cambios en la altura tonal se correlacionan con las estructuras de cabeza final, se predecía por tanto una segmentación de la lengua en la cual los elementos frecuentes ocupan una posición final en los bilingües L1euskera/L2castellano (esto es, en hablantes nativos de una lengua de cabeza final). Una segmentación de la lengua en la cual los elementos frecuentes ocupan una posición final en los hablantes bilingües L1castellano/L2euskera implicaría que estos bilingües habrían implementado el procedimiento de segmentación de su L2.

Los resultados del experimento C mostraron que la inserción de esta pista prosódica no conllevó preferencias de segmentación diferentes en ninguno de los dos grupos de bilingües, esto es, su preferencia de segmentación no difirió de la preferencia obtenida con una lengua artificial similar, aunque prosódicamente plana (experimento B). Adicionalmente, no se obtuvieron diferencias en las preferencias de segmentación entre los dos grupos de bilingües.

Una mayor preferencia por una segmentación en la cual los elementos frecuentes ocupan una posición final en el grupo de L1euskera pero no en el grupo de L1castellano, en comparación con las preferencias obtenidas en el experimento B, habría señalado que los bilingües de L1castellano no podían implementar la estrategia prosódica de segmentación de su L2. Una mayor segmentación en la cual los elementos frecuentes ocupan una posición final que en el experimento B en ambos grupos de bilingües habría sugerido que los bilingües L1castellano podían implementar las estrategias de segmentación de sus dos lenguas. Desafortunadamente, la ausencia de una preferencia de segmentación encontrada en ambos grupos de participantes no nos proporciona información sobre si los hablantes bilingües son capaces de implementar las estrategias de segmentación de sus L1 y L2.

Esta insensibilidad aparente hacia la pista prosódica en ambos grupos de bilingües nos impide extraer ninguna conclusión sobre las habilidades potenciales de los bilingües L1castellano/L2euskera para implementar la estrategia de segmentación de la L2. Las causas potenciales para esta insensibilidad hacia la pista prosódica se discutirán en la próxima sección.

6.2 Pistas estadísticas y prosódicas en la jerarquía de pistas segmentales

Uno de los objetivos de esta tesis doctoral era investigar el peso relativo que los hablantes adultos confieren a las pistas estadísticas y prosódicas, en comparación con otras pistas de segmentación. Los experimentos A y B se concentraban en el rol de las pistas estadísticas en la segmentación del habla y la interacción de pistas estadísticas y acústico-fonéticas, mientras que el experimento C investigaba el peso relativo dado por los hablantes adultos a las pistas prosódicas en comparación con las pistas estadísticas.

Como se describió en el capítulo 2, sección 2.3, las pistas estadísticas y prosódicas parecen pertenecer a los niveles más bajos en la jerarquía de pistas de segmentación de los hablantes adultos, propuesta por Mattys, White y Melhorn (2005). Tal y como han demostrado numerosos estudios que han puesto en competición pistas de segmentación, los hablantes adultos parecen apoyarse en la información léxica más que en cualquier otra pista de segmentación (Mattys et al., 2005). Cuando la información léxica no está disponible (por ejemplo en experimentos de aprendizaje de lenguas artificiales) los hablantes adultos confieren más peso a las pistas segmentales (pistas acústico-fonéticas y fonotácticas) que a las pistas prosódicas y estadísticas (Finn y Hudson Kam, 2008; Mattys et al., 2005, McQueen, 1998). Los hablantes adultos parecen hacer uso de la información estadística y prosódica como últimos recursos cuando ni la información léxica ni la segmental resultan informativas (Mattys, et al., 2005). Por último, las pistas prosódicas parecen superponerse a las pistas estadísticas, lo cual sugiere que las pistas estadísticas ocupan el nivel más bajo dentro de la jerarquía de pistas de segmentación (Langus, Marchetto, Bion y Nespor, 2012; Shukla, Nespor y Mehler, 2007).

El experimento A exploró la sensibilidad de los hablantes adultos hacia pistas estadísticas. En este experimento se obtuvo una preferencia general por una segmentación en la cual los elementos frecuentes ocupan una posición final en todos los grupos de participantes (bilingües de euskera/castellano y monolingües de castellano, inglés y francés). Argumenté que este sesgo podría haber sido causado por las propiedades acústico-fonéticas de la voz alemana empleada para sintetizar los estímulos. Concretamente, propuse que los valores de VOT de las oclusivas sordas del alemán podrían haber inducido a los participantes (ninguno de los cuales era hablante de alemán) a segmentar los estímulos en el inicio de las oclusivas sordas, las cuales aparecían siempre en el inicio de sílabas infrecuentes, resultando por tanto en una preferencia de segmentación de la lengua artificial en la cual los elementos frecuentes ocupan una posición final.

Esta propuesta se vio respaldada por la preferencia de segmentación en la cual los elementos frecuentes ocupan una posición inicial obtenida en el experimento B. La lengua artificial en el experimento B tenía una estructura similar a la lengua artificial en el experimento A, pero estaba sintetizada a una voz en castellano, esto es, una voz que contenía fonemas comunes a los repertorios de las lenguas nativas de los participantes (euskera, castellano), eliminando por tanto las pistas acústico-fonéticas propuestas como la causa de la segmentación general en la cual los elementos frecuentes ocupan una posición final obtenida en el experimento A. Este cambio en la preferencia de segmentación mostrado por todos los grupos en el experimento B, en comparación con la obtenida en el experimento A, sugiere que las pistas acústico-fonéticas presentes en el input, y no la información estadística, determinaron la preferencia de segmentación de los participantes. Este resultado coincide con la prevalencia de pistas acústico-fonéticas (coarticulación) sobre pistas estadísticas obtenida por Fernandes, Ventura y Kolinsky (2007) con hablantes adultos y proporciona evidencia que apoya la concepción jerárquica de las pistas de segmentación propuesta por Mattys et al. (2005).

Aunque, como se mencionó anteriormente, todos los participantes en el experimento A prefirieron una segmentación general de la lengua en la cual los elementos frecuentes ocupan una posición final, se obtuvieron diferencias en las preferencias de segmentación entre los participantes hablantes nativos de lenguas OV (bilingües L1euskera) y los hablantes nativos de lenguas VO (monolingües de castellano, francés e inglés). Esta diferencia cumplió las predicciones de la *frequency-based bootstrapping hypothesis* —la cual será tratada en la próxima sección—y sugiere que, aunque principalmente dirigidos por las propiedades acústico-fonéticas del input, la información estadística moduló empero las preferencias de segmentación de los participantes, coincidiendo así con los resultados de numerosos estudios que han mostrado que la segmentación está modulada por las pistas de segmentación pertenecientes a los niveles inferiores de la jerarquía, disponibles para los hablantes adultos (Andruski, Blumstein y Burton, 1994; Dumay, Content y Frauenfelder, 1999; McQueen, Norris y Cutler, 1994; Norris, McQueen y Cutler, 1995; Sanders y Neville, 2000; Smith y Hawkins, 2000; Vroomen, van Zon y de Gelder, 1996).

En el experimento C, los participantes escucharon una lengua artificial que consistía en una alternancia de sílabas frecuentes e infrecuentes similar a la lengua en el experimento B y que por tanto contenía la misma información estadística. Además de esta pista estadística, la lengua artificial en el experimento C contenía una pista prosódica para la segmentación de la lengua: la alternancia de subidas y bajadas tonales. Como se mencionó en la sección previa, Bion et al. (2011) y Nespor et al. (2008) proponen que esta pista prosódica lleva a la segmentación de una cadena de sonidos en la que la prominencia ocurre en posición inicial, esto es, una segmentación de la cadena del experimento C en la cual los elementos frecuentes ocupan una posición final. Tanto la información estadística como la información prosódica deberían indicar una segmentación de la lengua en la cual los elementos frecuentes ocupan una posición final a hablantes de lenguas OV. Por el contrario, estas dos pistas proporcionarían información contradictoria a hablantes de lenguas VO: la pista estadística señalaría una segmentación en la cual los elementos frecuentes ocupan una posición inicial, mientras que la información prosódica indicaría una segmentación de la lengua en la cual los elementos frecuentes ocupan una posición final. La existencia de diferencias en las preferencias de segmentación de los bilingües L1euskera y L1castellano revelaría que los participantes habían integrado estas dos pistas. Sin embargo, no se obtuvo esta diferencia. Más aún, como se mencionó en la sección previa, la información prosódica no influyó significativamente en las preferencias de segmentación de ninguno de los grupos de participantes.

Estos resultados sugieren que los participantes no se apoyaron en las pistas prosódicas para segmentar el input. En su lugar, los participantes mostraron una preferencia de segmentación en la cual los elementos frecuentes ocupan una posición inicial similar a la obtenida en el experimento B. En el experimento B argumenté que el sesgo general hacia una segmentación en la cual los elementos frecuentes ocupan una posición inicial encontrada en todos los grupos podría resultar de potenciales pistas acústico-fonéticas presentes en el input. Dado que las lenguas artificiales en los experimentos B y C contenían la misma información acústico-fonética, la aparente insensibilidad hacia la pista prosódica observada en el experimento C sugeriría que la pista prosódica fue sobrepasada por la información segmental, tal y como predicen Mattys et al. (2005). En cualquier caso, es imposible determinar cuál de las dos situaciones siguientes tuvo lugar: (a) si la integración de estas dos pistas, ambas situadas en un nivel más bajo en la jerarquía de pistas que la información segmental, puede no haber bastado para sobrepasar la potencial información segmental presente en el input, o (b) si los participantes no fueron capaces de integrar las dos pistas de segmentación. Esta indeterminación abre la puerta a futuras investigaciones.

En el capítulo 4 propuse una explicación alternativa para el sesgo general hacia una segmentación en la cual los elementos frecuentes ocupan una posición inicial encontrado en el experimento B, esto es, que este sesgo podría revelar una preferencia por defecto por la segmentación en la cual los elementos frecuentes ocupan una posición inicial, la cual sería preferida universalmente. En este escenario, la preferencia de segmentación en la cual los elementos frecuentes ocupan una posición inicial encontrada en el experimento C sugeriría que ni la información prosódica ni la información convergente proporcionada por pistas prosódicas y estadísticas fueron lo suficientemente poderosas como para sobrepasar esta preferencia por defecto.

Aunque se requiere más investigación para poder determinar si la preferencia de segmentación en la cual los elementos frecuentes ocupan una posición inicial encontrada en los experimentos B y C resulta de la información acústico-fonética presente en el input o revela en cambio una estrategia de segmentación preferida universalmente, la ausencia de un efecto causado por la pista prosódica introducida en el experimento C revela que la información prosódica juega un papel menor en la segmentación del habla en hablantes adultos.

6.3 Disponibilidad para los adultos de pistas basadas en la frecuencia y pistas prosódicas como mecanismos para emprender la adquisición de la sintaxis en la segmentación del habla

El tercer objetivo de los experimentos de la presente tesis doctoral era examinar si los hablantes adultos emplean pistas estadísticas y prosódicas en la segmentación del habla. Concretamente, la pista estadística investigada era la distribución de la frecuencia y el orden relativo de elementos frecuentes e infrecuentes, reflejando así la distribución de la frecuencia de los elementos funcionales y de contenido en las lenguas naturales. La pista prosódica examinada era la realización acústica y la localización relativa de la prominencia dentro de los sintagmas fonológicos. Se ha propuesto que estas dos pistas podrían jugar un rol crucial para emprender la adquisición de la sintaxis, permitiendo a los niños establecer el valor del parámetro de direccionalidad de la cabeza sintáctica antes de tener conocimiento léxico, dada la correlación existente entre estas dos pistas y el orden de cabezas y complementos en las lenguas naturales (Christophe, Guasti, Nespor, Dupoux y van Ooyen, 1997; Gervain, 2007; Nespor, Guasti y Christophe, 1996; Nespor, Shukla, van de Vijver, Avesani, Schraudolf y Donati, 2008).

Como se presentó en el capítulo 2, sección 2.1.2, los elementos funcionales son elementos altamente frecuentes en las lenguas naturales, que tienden a aparecer en el inicio de los sintagmas en las lenguas de cabeza inicial (Gervain, 2007). La *frequency-based bootstrapping hypothesis* (Gervain, 2007) propone que los niños computan la distribución de la frecuencia de los elementos funcionales y su orden de aparición relativo a las palabras de contenido, construyendo así una representación rudimentaria del orden de cabezas y complementos en la lengua que se está adquiriendo.

Como se presentó en el capítulo 3, Gervain (2007) llevó a cabo una serie de experimentos sobre aprendizaje de lenguas artificiales con niños y hablantes adultos de varias lenguas de cabeza inicial (VO) y cabeza final (OV) y encontró que tanto los adultos como los niños eran capaces de computar la distribución de la frecuencia de los elementos en una lengua artificial ambigua, y que sus preferencias de segmentación parecían correlacionarse con el orden relativo de los elementos funcionales (elementos frecuentes) y palabras de contenido (elementos infrecuentes) en sus lenguas nativas. Así, los niños de 8 meses de edad que estaban aprendiendo una lengua de cabeza final (japonés) miraron más prolongadamente los estímulos con un orden en el cual los elementos frecuentes ocupan una posición final que los estímulos con un orden en el cual los elementos frecuentes ocupan una posición inicial, mientras que los niños que estaban aprendiendo una lengua de cabeza inicial (italiano) mostraron el patrón opuesto. Igualmente, los hablantes adultos de lenguas de cabeza final (euskera, japonés, húngaro) eligieron una segmentación de la lengua en la cual los elementos frecuentes ocupan una posición final, mientras que los hablantes de lenguas de cabeza inicial (francés, italiano) eligieron una segmentación de la lengua en la cual los elementos frecuentes ocupan una posición inicial.

El experimento A de la presente tesis doctoral tenía como objetivo explorar la disponibilidad de las estrategias de segmentación basadas en la frecuencia para los hablantes adultos de lenguas de cabeza inicial y cabeza final. Para este fin, se presentó una lengua artificial ambigua similar a la creada y empleada por Gervain (2007)[39] a dos grupos de hablantes nativos de dos lenguas previamente no examinadas (inglés y castellano, ambas lenguas VO de cabeza inicial), así como a hablantes nativos de dos lenguas ya examinadas por Gervain (euskera, OV y francés, VO).

Tal y como predice la *frequency-based bootstrapping hypothesis*, los hablantes de una lengua OV (euskera) eligieron una segmentación de la lengua en la cual los elementos frecuentes ocupan una posición final más a menudo que los hablantes de las lenguas VO (francés, castellano e inglés). Este resultado sugiere que los hablantes adultos computaron la pista de frecuencia contenida en el input y que los hablantes de lenguas VO y OV tenían diferentes representaciones del orden relativo de elementos frecuentes e infrecuentes, en correlación con el orden relativo de elementos funcionales y de contenido de sus lenguas nativas. Estas diferentes preferencias de segmentación replicaron por tanto los resultados obtenidos por Gervain (2007) y proporcionaron evidencia que apoyaba la propuesta de que la distribución de la frecuencia de los elementos en el input constituye una pista útil para la segmentación del habla y que los elementos frecuentes en las lenguas naturales (elementos funcionales) actúan como puntos de anclaje para la estructura sintáctica.

[39] Idénticas en estructura, léxico, familiarización y estímulos experimentales, pero sintetizados a una voz alemana en lugar de la voz española usada en Gervain (2007).

No obstante, los hablantes de lenguas VO en el experimento A no mostraron la preferencia de segmentación en la cual los elementos frecuentes ocupan una posición inicial encontrada por Gervain (2007), sino una preferencia de segmentación en la cual los elementos frecuentes ocupan una posición final, aunque menor que la preferencia observada en hablantes de lenguas OV, como resultado del sesgo causado por las propiedades acústico-fonéticas de la voz alemana empleada para sintetizar los estímulos—sesgo discutido en la sección anterior. Las pistas acústico-fonéticas se superpusieron por tanto a las pistas de frecuencia presentes en el input.

Como se presentó en el capítulo 2, sección 2.3, los niños acceden a la jerarquía de pistas segmentales desde el nivel inferior y en dirección ascendente, a diferencia de los hablantes adultos, quienes acceden a la jerarquía desde el nivel superior. En el comienzo del proceso de adquisición de la lengua, los niños se apoyan en información estadística y prosódica, previamente a la adquisición de pistas segmentales y conocimiento léxico (Johnson y Tyler, 2010; Jusczyk, Houston y Newsome, 1999). Por tanto, el hecho de que las pistas de segmentación basadas en la frecuencia modulen (aunque no determinen) las preferencias de segmentación de los hablantes adultos, sumado a las propiedades salientes de los elementos funcionales en las lenguas naturales y el rol crucial que la información estadística parece jugar al comienzo de la adquisición de la lengua, dan plausibilidad a la propuesta de Gervain (2007) de que las pistas basadas en la frecuencia podrían permitir a los niños emprender la adquisición de un parámetro sintáctico principal, esto es, el parámetro de direccionalidad de la cabeza sintáctica.

En el experimento B, hablantes nativos de euskera (OV) y castellano (VO) escucharon a una lengua artificial con la misma estructura que la lengua empleada en el experimento A y en Gervain (2007) pero sintetizada a la misma voz española empleada en los experimentos en Gervain (2007), en lugar de la voz alemana empleada en el experimento A. Este cambio en la voz pretendía examinar si los hablantes de castellano (VO) mostraban la segmentación en la cual los elementos frecuentes ocupan una posición inicial predicha por la *frequency-based bootstrapping hypothesis*, una vez eliminada la información acústica-fonética causante del sesgo general hacia una segmentación en la cual los elementos frecuentes ocupan una posición final observado en el experimento A.

Cumpliendo las predicciones de esta hipótesis, los hablantes nativos de castellano (monolingües de castellano y bilingües L1castellano) prefirieron una segmentación en la cual los elementos frecuentes ocupan una posición inicial de la lengua. Sorprendentemente, los participantes L1euskera mostraron una preferencia similar por una segmentación en la cual los elementos frecuentes ocupan una posición inicial, opuesta a la preferencia de segmentación en la cual los elementos frecuentes ocupan una posición final predicha por esta hipótesis y obtenida con una población similar en Gervain (2007) y en el experimento A. Como se discutió en el capítulo 4, esta preferencia de segmentación en la cual los elementos frecuentes ocupan una posición inicial es sorprendente, dado que la lengua artificial en el experimento B tenía la misma estructura y estaba sintetizada a la misma voz que la lengua artificial empleada en Gervain (2007).

Como se mencionó en la sección anterior, propuse dos explicaciones tentativas para la preferencia de segmentación general obtenida en la cual los elementos frecuentes ocupan una posición inicial, a saber, la influencia de pistas acústico-fonéticas y la posibilidad de que la segmentación en la cual los elementos frecuentes ocupan una posición inicial sea la segmentación preferida universalmente. El léxico de la lengua artificial en el experimento B contenía dos fonemas que no aparecían en Gervain (2007), ni en el léxico del experimento A: /*l*/ y /*s*/[40]. Sin embargo, parece poco probable que estos dos segmentos, los cuales forman parte de los repertorios fonológicos tanto del euskera como del castellano, hayan causado este sesgo. Esta posibilidad se descartaría totalmente mediante un experimento de control en el cual el léxico sólo contuviera exactamente los mismos fonemas que los de la lengua artificial de Gervain (2007). Por otra parte, la segmentación general en la cual los elementos frecuentes ocupan una posición inicial encontrada en los experimentos B y C proporciona evidencia que apoya la propuesta de que la segmentación en la cual los elementos frecuentes ocupan una posición inicial podría ser el procedimiento de segmentación por defecto. No obstante, esta propuesta no puede explicar la preferencia de segmentación en la cual los elementos frecuentes ocupan una posición final obtenida en Gervain (2007) con hablantes de japonés y euskera (OV).

[40] Estos fonemas se incluyeron en sustitución de otros fonemas presentes en el léxico en el experimento A, para evitar efectos potenciales fonotácticos y alofónicos (tal y como se describió en la metodología del experimento B en el capítulo 4).

Por tanto, no puedo dar cuenta de forma satisfactoria de la causa de la preferencia general de segmentación obtenida en la cual los elementos frecuentes ocupan una posición inicial, o más bien, de la ausencia de una segmentación en la cual los elementos frecuentes ocupan una posición final en hablantes nativos de euskera. Quedan pendientes futuras investigaciones que tratarán de clarificar esta cuestión.

El experimento C tenía como objetivo investigar una pista prosódica que varios autores que trabajan en el marco de la *phonological bootstrapping hypothesis* han propuesto que puede ayudar a los niños a establecer el valor del parámetro de direccionalidad de la cabeza sintáctica. Esta pista es la localización y realización de la prominencia principal en los sintagmas fonológicos (Christophe, Guasti, Nespor, Dupoux y van Ooyen, 1997; Nespor, Guasti y Christophe, 1996; Nespor, Shukla, van de Vijver, Avesani, Schraudolf y Donati, 2008). Concretamente, el experimento C pretendía examinar si los hablantes bilingües de euskera/castellano demostraban sensibilidad hacia una pista acústica asociada al orden de palabras característico de una lengua de cabeza final como el euskera.

Tal y como se presentó en el capítulo 2, sección 2.2, tanto los adultos como los niños son sensibles a la información prosódica contenida en el input. La sensibilidad de los niños hacia las pistas prosódicas emerge en los estadios iniciales del desarrollo. Estas pistas podrían por tanto ser cruciales para emprender la adquisición del léxico y la sintaxis. La localización de la prominencia principal dentro de los sintagmas fonológicos se correlaciona con el orden relativo de cabezas y complementos en las lenguas naturales (Nespor y Vogel, 1986) y se ha propuesto que esta prominencia tiene diferentes realizaciones acústicas en estructuras de cabeza inicial y cabeza final (Nespor et al., 2008), según la ley Yámbico/Trocaico (Hayes, 1995). En las lenguas de cabeza inicial la prominencia recae en la palabra que se encuentra situada en el límite derecho del sintagma fonológico y se realiza mediante a un aumento de la duración, mientras que en las lenguas de cabeza final la prominencia principal recae en la palabra que se encuentra situada en el límite izquierdo y se realiza mediante una subida de la altura tonal (Nespor y Vogel, 1986; Nespor et al., 2008). Bion, Benavides-Varela y Nespor (2011) mostraron que los hablantes adultos agrupaban estímulos que contenían cambios en la altura tonal en grupos de prominencia inicial, y estímulos que contenían cambios en la duración en grupos con prominencia final.

Para examinar la sensibilidad de los hablantes bilingües hacia esta pista presenté a participantes bilingües de euskera/castellano con una lengua artificial similar a la empleada en el experimento B, con la manipulación añadida de una subida tonal en cada sílaba infrecuente, señalando por tanto una segmentación de la lengua en la cual los elementos frecuentes ocupan una posición final. No obstante, ningún grupo de participantes eligió la segmentación esperada en la cual los elementos frecuentes ocupan una posición final. Por tanto, los resultados obtenidos en el experimento C no replicaron la agrupación caracterizada por una prominencia inicial de estímulos que contenían cambios en altura tonal obtenida en Bion et al. (2011). Hay que apuntar, no obstante, que a diferencia de Bion et al. (2011) la lengua artificial en el experimento C contenía adicionalmente una pista estadística.

Los experimentos de la presente tesis doctoral muestran así que las estrategias de segmentación basadas en la frecuencia que Gervain (2007) propone que podrían ayudar a los niños a emprender la adquisición de la sintaxis también están disponibles para los hablantes adultos, aunque son sobrepasadas con facilidad por otras pistas presentes en el input. Por otra parte, los resultados obtenidos en el experimento C muestran que los hablantes adultos no usan la pista prosódica que Nespor et al. (2008) proponen que podría asistir a los niños a emprender la adquisición de la sintaxis. La influencia de estas dos pistas parece por tanto ser limitada en la segmentación de habla por hablantes adultos. No obstante, como se describió antes, las pistas estadísticas y prosódicas parecen jugar un rol crucial en los estadios iniciales de adquisición de la lengua. Estas dos pistas son las primeras detectadas por los niños, quienes comienzan a integrar la información prosódica y estadística sobre los 8 ó 9 meses de edad (Johnson y Tyler, 2010; Jusczyk, Houston y Newsome, 1999; Morgan, 1994; Morgan y Saffran, 1995).

De hecho, se ha propuesto que la adquisición de la sintaxis podría emprenderse como resultado de la integración de estas dos fuentes de información, esto es, pistas estadísticas y prosódicas (Gervain, 2007; Shukla y Nespor, 2010). Christophe, Millotte, Bernal y Lidz (2008) proponen que la prosodia a nivel de sintagma y los elementos funcionales podrían permitir a los niños llevar a cabo un análisis inicial del input, emprendiendo la adquisición de estructura sintáctica y categorías sintácticas. Según Christophe et al. (2008), la información prosódica asistiría a los niños en la localización de los límites de los sintagmas, mientras que los elementos funcionales les ayudarían a inferir la categoría sintáctica de las palabras adyacentes, dentro de los constituyentes sintácticos.

Como se presentó en el capítulo 2, los niños de 6 a 9 meses de edad muestran sensibilidad hacia los marcadores prosódicos de los sintagmas fonológicos, esto es, los niños prefieren escuchar sintagmas prosódicamente bien formados a sintagmas mal formados (Jusczyk, et al, 1992; Soderstrom et al., 2003). Además, la segmentación de las palabras está constreñida por los dominios de los sintagmas fonológicos. Los niños no consideran secuencias que se extienden a través de los límites de sintagmas fonológicos como candidatos a constituir palabras (Gout, Christophe y Morgan, 2004). Igualmente, la sensibilidad hacia los elementos funcionales surge muy pronto en el desarrollo, debido a sus salientes propiedades estadísticas, fonológicas y distribucionales. A los 8 meses de edad, los niños no sólo segmentan los elementos funcionales del habla fluida y tienen una representación detallada de al menos los elementos funcionales de mayor frecuencia, sino que también hacen uso de estos elementos funcionales para segmentar palabras de contenido adyacentes (Höhle y Weissenborn, 2003; Shi, Cutler, Werker y Cruishank, 2006; Shi, Marquis y Gauthier, 2006). A los 24 meses de edad los niños han reunido un conocimiento parcial sobre los patrones de coocurrencia entre las categorías de los elementos funcionales y las palabras de contenido (Gerken y McIntosh, 1993). Los elementos funcionales, los cuales tienden a aparecer en los extremos de los sintagmas, han sido propuestos como marcadores que ayudan a aprender las reglas sintácticas de las lenguas (Green, 1979; Valian y Coulson, 1988).

Por tanto, es plausible que la información conjunta proporcionada por los elementos funcionales y la prosodia a nivel de sintagma pueda permitir a los niños emprender la adquisición de la sintaxis. La prosodia a nivel de sintagma permitiría a los niños dividir el input en pedazos, caracterizados por la presencia de elementos funcionales en los extremos de estos pedazos más pequeños. Las propiedades distribucionales de los elementos funcionales podrían a su vez permitir a los niños descubrir los patrones de coocurrencia de categorías específicas de elementos funcionales y de palabras de contenido.

De hecho, Endress y Mehler (2009) proponen que los extremos son primitivos perceptuales que ayudan a los oyentes a codificar la información. Estudios sobre el aprendizaje de lenguas artificiales ha revelado que los participantes adultos aprenden con éxito una dependencia no adyacente cuando la dependencia aparece en los extremos de una secuencia, pero no la aprenden cuando la dependencia aparece en una posición interna dentro de la secuencia (Endress y Mehler, 2009). Igualmente, los hablantes adultos aprenden nuevas restricciones fonotácticas cuando estas restricciones aparecen en los extremos (C_1VccVC_2), pero no cuando los segmentos críticos aparecen en posición interna en la palabra (cVC_1C_2Vc), a menos que se les proporcione una pista adicional (por ejemplo que los conjuntos de consonantes permitidas en *C1* y *C2* pertenezcan a clases naturales diferentes, Endress y Mehler, 2010). Además, los hablantes adultos generalizan estructuras basadas en repeticiones cuando estas repeticiones aparecen en los extremos de las secuencias (ABCDE*FF*), pero no cuando aparecen en posición interna en las secuencias (ABC*DD*EF) (Endress, Scholl y Mehler, 2005). Más aún, como se describió en el capítulo 2, la prominencia principal recae en los extremos de los sintagmas fonológicos en las lenguas naturales y, como mencionan Endress y Mehler (2009), en las lenguas naturales los afijos aparecen generalmente en los extremos de las palabras (sufijos, prefijos) en lugar de en posición interna. Por último, diversos estudios sobre las estrategias que las madres emplean para enseñar palabras nuevas a los niños han revelado que las madres tienden a situar la información nueva en el extremo final de las emisiones (Woodward y Aslin, 1990, 1991, 1992, citados en Aslin, Woodward, LaMendola y Bever, 1996).

6.4 Conclusiones

→ Los hablantes bilingües pueden emplear las estrategias de segmentación basadas en la frecuencia características de sus dos lenguas, aunque la adquisición de la estrategia de segmentación de la L2 parece estar limitada. Queda pendiente determinar las causas de esta limitación.

→ Las pistas estadísticas y prosódicas parecen jugar un rol menor en la segmentación del habla de los hablantes adultos. Otras pistas tales como las pistas acústico-fonéticas se superponen a las estadísticas y prosódicas, apoyando así una concepción jerárquica de las pistas de segmentación según la cual la información estadística y la prosódica son las pistas de segmentación a las que los hablantes adultos atribuyen menor peso.

→ La segmentación en la cual los elementos frecuentes ocupan una posición inicial podría ser la estrategia de segmentación universalmente preferida. Futuras investigaciones se ocuparán de examinar esta hipótesis.

→ Las estrategias de segmentación basadas en la frecuencia son pistas disponibles para los hablantes adultos en la segmentación del habla. Así, los hablantes adultos computan la distribución de la frecuencia de los elementos, y el orden relativo de los elementos frecuentes (elementos funcionales) e infrecuentes (palabras de contenido) en sus lenguas nativas modula las preferencias de segmentación de los hablantes.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh, UK: Edinburgh University Press.

Altenberg, E. P., & Cairns, H. S. (1983). The effects of phonotactic constraints on lexical processing in bilingual and monolingual subjects. *Journal of Verbal Learning and Verbal Behavior, 22*, 174–188.

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition, 52*, 163–187.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9* (4), 321–324.

Aslin, R. N., Woodward, J. Z., LaMendola, N. P., & Bever, T. G. (1996). Models of word segmentation in fluent maternal speech to infants. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to syntax in early acquisition* (pp. 117–134). Hillsdale, USA: Erlbaum.

Bagou, O., Fougeron, C., & Frauenfelder, U. H. (2002). Contribution of prosody to the segmentation and storage of "words" in the acquisition of a new mini-language. In Proceedings from *5th Speech Production Workshop: Models and Data*. Germany: Kloster Seeon.

Baker, M. (2001). *The atoms of language: The mind's hidden rules of grammar.* New York, USA: Basic Books.

Bion, R. A. H., Benavides-Varela, S., & Nespor, M. (2011). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech, 54* (1), 123–140.

Bloom, L. (1970). *Language development: Form and function in emerging grammars.* Cambridge, USA: MIT Press.

Boersma, P. & Weenink, D. (2012). Praat: Doing phonetics by computer [Computer program]. Version 5.3.19, retrieved 24 June 2012 from http://www.praat.org/

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: the role of consonants and vowels in continuous speech processing. *Psychological Science, 16* (6), 451–459.

Bosch, L., & Sebastián-Gallés, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition, 65*, 33–69.

Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy, 2* (1), 29–49.

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech, 46* (2/3), 217–243.

Bradley, D. C., Sánchez-Casas, R. M., & García-Albea, J. E. (1993). The status of the syllable in the perception of Spanish and English. *Language and Cognitive Processes, 8* (2), 197–233.

Braunschweiler, N. (1997). Integrated cues of voicing and vowel length in German: A production study. *Language and Speech, 40* (4), 353–376.

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 61*, 93–125.

Brown, R. (1973). *A first language.* Cambridge, USA: Harvard University Press.

Butterfied, S., & Cutler, A. (1988). Segmentation errors by human listeners: Evidence for a prosodic segmentation strategy. In W. Ainsworth, & J. Holmes (Eds.), Proceedings from *SPEECH'88: 7th Symposium of the Federation of Acoustic Societies of Europe, Vol. 3* (pp. 827–833). Edinburgh, UK: Institute of Acoustics.

Cairns, P., Shillcock, R., Chater, N., & Levy, J. (1994). Modelling the acquisition of lexical segmentation. In Proceedings from *26th Child Language Research Forum, 1994* (pp. 32–41). CSLI, USA: University of Chicago Press.

Cairns, P., Shillcock, R., Chater, N., & Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology, 33*, 111–153.

Caramazza, A., & Yeni-Komshian, G. H. (1974). Voice onset time in two French dialects. *Journal of Phonetics, 2*, 239–245.

Castañeda, M. L. (1986). El V.O. T. de las oclusivas sordas y sonoras españolas. *Estudios de fonética experimental II*, 91–108.

Chambers, K. E., Onishi, K. H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition, 87*, B69–B77.

Chomsky, N. (1981). *Lectures on government and binding.* Dordrecht, The Netherlands: Foris.

Chomsky, N. (1995). *The minimalist program.* Cambridge, USA: MIT Press.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English.* New York, USA: Harper & Row.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes, 13* (2/3), 221–268.

Christiansen, M. H., Conway, C, M., & Curtin, S. (2005). Multiple-cue integration in language acquisition: A connectionist model of speech segmentation and rule-like behavior. In J. W. Minett and W. S.-Y. Wang (Eds.), *Language acquisition, change and emergence: Essays in evolutionary linguistics* (pp. 205–249) Hong Kong: City University of Hong Kong Press.

Christophe, A., & Dupoux, E. (1996). Bootstrapping lexical acquisition: The role of prosodic structure. *The Linguistic Review, 13*, 383–412.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America, 95* (3), 1570–1580.

Christophe, A., Guasti, M. T., Nespor, M., Dupoux, E., & van Ooyen, B. (1997). Reflections on phonological bootstrapping: Its role for lexical and syntactic acquisition. *Language and Cognitive Processes, 12* (5/6), 585–612.

Christophe, A., Mehler, J., & Sebastián-Gallés, N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy, 2* (3), 385–394.

Christophe, A., Millotte, S., Bernal, S., & Lidz, J. (2008). Bootstrapping lexical and syntactic acquisition. *Language and Speech, 51* (1/2), 61–75.

Christophe, A., & Morton, J. (1998). Is Dutch native English? Linguistic analyses by 2-month-olds. *Developmental Science, 1* (2), 215–219.

Christophe, A., Nespor, M., Guasti, M. T., & van Ooyen, B. (2003). Prosodic structure and syntactic acquisition: the case of the head-direction parameter. *Developmental Science, 6* (2), 211–220.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access I. Adult data. *Journal of Memory and Language*, *51*, 523–547.

Church, K. W. (1987). Phonological parsing and lexical retrieval. *Cognition*, *25*, 53–69.

Cinque, G. (1993). A null theory of phrase and compound stress. *Linguistic inquiry*, *24* (2), 239–297.

Cole, R. A., & Jakimik, J. (1980). A model of speech perception. In R.A. Cole (Ed.), *Perception and production of fluent speech* (pp. 133–163). Hillsdale, USA: Erlbaum.

Colomé, A. (2001). Lexical activation in bilinguals' speech production: Language-Sspecific or language-independent? *Journal of Memory and Language*, *45*, 725–736.

Condon, W. S., & Sander, L. W. (1974). Neonate movement is synchronized with adult speech: Interactional participation and language acquisition. *Science*, *183* (4120), 99–101.

Cooper, G., & Meyer, L. (1960). *The rhythmic structure of music*. Chicago, USA: University of Chicago Press.

Cooper, W., & Paccia-Cooper, J. (1980). *Syntax and speech.* Cambridge, USA: Harvard University Press.

Costa, A., Caramazza, A., & Sebastián-Gallés, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26* (5), 1283–1296.

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, *50*, 491–511.

Costa, A., Santesteban, M., & Ivanova, I. (2006). How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32* (5), 1057–1074.

Cruttenden, A. (1986). *Intonation*. Cambridge, UK: Cambridge University Press.

Curtin, S., Mintz, T. H., & Byrd, D. (2001). Coarticulatory cues enhance infants' recognition of syllable sequences in speech. In A. H.-J. Do, L. Domínguez, & A. Johansen (Eds.), Proceedings from *BUCLD: 25th Annual Boston University Conference on Language Development, Vol. 1* (pp. 190–201). Somerville, USA: Cascadilla Press.

Cutler, A. (1993). Phonological cues to open- and closed-class words in the processing of spoken sentences. *Journal of Psycholinguistic Research, 22* (2), 109–131.

Cutler, A. (2001). Listening to a second language through the ears of a first. *Interpreting, 5* (1), 1–23.

Cutler, A. (2004). Segmentation of spoken language by normal adult listeners. In R. D. Kent (Ed.), *The MIT encyclopedia of communication sciences and disorders* (392–395). Cambridge, USA: MIT Press.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language, 31*, 218–236.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2*, 133–142.

Cutler, A., Hawkins, J. A., & Gilligan, G. (1985). The suffixing preference: a processing explanation. *Linguistics, 23*, 723–758.

Cutler, A., Mehler, J., Norris, D., & Seguí, J. (1983). A language-specific comprehension strategy. *Nature, 304* (5922), 159–160.

Cutler, A., Mehler, J., Norris, D., & Seguí, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, *25*, 385–400.

Cutler, A., Mehler, J., Norris, D., & Seguí, J. (1989). Limits on bilingualism. *Nature*, *340* (6230), 229–230.

Cutler, A., Mehler, J., Norris, D., & Seguí, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, *24*, 381–410.

Cutler, A., Murty, L., & Otake, T. (2003). Rhythmic similarity effects in non-native listening? In M. Solé, D. Recasens & J. Romero (Eds.), Proceedings from *15th International Congress of Phonetic Sciences, Vol. 1* (pp.329–332). Adelaida, Australia: Causal Productions.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14* (1), 113–121.

Cutler, A., & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language, 33*, 824–844.

Dasher, R., & Bolinger, D. (1982). On pre-accentual lengthening. *Journal of the International Phonetic Association, 12*, 58–69.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics, 11* (1), 51–62.

Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience, 12* (4), 635–647.

Dehaene-Lambertz, G., & Houston, D. (1998). Faster orientation latencies toward native language in two-month-old infants. *Language and Speech, 41* (1), 21–43.

Dumay, N., Content, A., & Frauenfelder, U. H. (1999). Acoustic-phonetic cues to word boundary location: evidence from word spotting. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A. C. Bailey (Eds.), Proceedings from *14th International Congress of Phonetic Sciences* (pp. 281–284). San Francisco, USA: University of California.

Dumay, N., Frauenfelder, U. H., & Content, A. (2002). The role of the syllable in lexical segmentation in French: Word-spotting data. *Brain and Language, 81*, 144–161.

Duñabeitia, J. A., Cholin, J., Corral, J., Perea, M., & Carreiras, M. (2010). SYLLABARIUM: An online application for deriving complete statistics for Basque and Spanish syllables. *Behavior Research Methods, 42*, 118–125.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance, 25* (6), 1568–1578.

Dutoit, T. (1997). *An introduction to text-to-speech synthesis.* Dordrecht,The Netherlands: Kluwer.

Endress, A. D., & Mehler, J. (2009). Primitive computations in speech processing. *The Quarterly Journal of Experimental Psychology, 62* (11), 2187–2209.

Endress, A. D., & Mehler, J. (2010). Perceptual constraints in phonotactic learning. *Journal of Experimental Psychology: Human Perception and Performance, 36* (1), 235–250.

Endress, A. D., Scholl, B. J., & Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology: General, 134* (3), 406–419.

Etxebarria, M. (2004). Español y euskera en contacto: Resultados lingüísticos. *Revista Internacional de Lingüística Iberoamericana II, 2* (4), 131–145.

Fais, L., Kajikawa, S., Werker, J., & Amano, S. (2005). Japanese listeners' perceptions of phonotactic violations. *Language and Speech, 48* (2), 185–201.

Fernandes, T., Ventura, P., & Kolinsky, R. (2007). Statistical information and coarticulation as cues to word boundaries: A matter of signal quality. *Perception & Psychophysics, 69* (6), 856–864.

Finn, A. S., & Hudson Kam, C. L. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition, 108*, 477–499.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12* (6), 499–504.

Fisher, C., & Tokura, H. (1996). Prosody in speech to infants: Direct and indirect acoustic cues to syntactic structure. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to syntax in early acquisition* (pp. 343–363). Hillsdale, USA: Erlbaum.

Flege, J. E., & Wang, C. (1989). Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/-/d/ contrast. *Journal of Phonetics, 17*, 299–315.

Fowler, C. A., Best, C. T., & McRoberts, G. W. (1990). Young infants' perception of liquid coarticulatory influences on following stop consonants. *Perception & Psychophysics, 48* (6), 559–570.

Friederici, A. D., & Wessels, J. M. I. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception & Psychophysics, 54* (3), 287–295.

Friedrich, M., & Friederici, A. D. (2005). Phonotactic knowledge and lexical-semantic processing in one-year-olds: Brain responses to words and nonsense words in picture contexts. *Journal of Cognitive Neuroscience, 17* (11), 1785–1802.

Fudge, E. (1984). *English word-stress*. London, UK: George Allen & Unwin.

Gambell, T., & Yang, C. (2004). Statistics learning and universal grammar: Modelling word segmentation. In W. G. Sakas (Ed.), Proceedings from *20th International Conference on Computational Linguistics* (pp. 51–54). Geneva, Switzerland.

Gerken, L., Jusczyk, P. W., & Mandel, D. R. (1994). When prosody fails to cue syntactic structure: 9-month-olds' sensitivity to phonological versus syntactic phrases. *Cognition, 51*, 237–265.

Gerken, L., Landau, B., & Remez, R. E. (1990). Function morphemes in young children's speech perception and production. *Developmental Psychology, 26* (2), 204–216.

Gerken, L., & McIntosh, B. J. (1993). Interplay of function morphemes and prosody in early language. *Developmental Psychology, 29* (3), 448–457.

Gervain, J. (2007). *From the richness of the signal to the poverty of the stimulus: Early mechanisms of language acquisition* (Unpublished doctoral dissertation). SISSA, Trieste, Italy.

Gervain, J., Sebastián-Gallés, N., Díaz, B., Laka, I., Mazuka, R., Yamane, N., Nespor, M., & Mehler, J. (submitted). Word frequency bootstraps word order: Cross-linguistic evidence.

Gleitman, L. R., & Wanner, E. (1982). Language acquisition: The state of the state of the art. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 3–48). Cambridge, UK: Cambridge University Press.

Golinkoff, R. M., Hirsh-Pasek, K., & Schweisguth, M. A. (2001). A reappraisal of young children's knowledge of grammatical morphemes. In J. Weissenborn & B. Höhle (Eds.), *Approaches to bootstrapping: Phonological, syntactic and neurophysiological aspects of early language acquisition, Vol. 1* (pp. 176–188). Amsterdam, The Netherlands: John Benjamins.

Golinkoff, R. M., Schweisguth, M. A., & Hirsh-Pasek, K. (1992). Young children can use linguistic information to assign novel words to syntactic categories. Paper presented at *International Conference on Infant Studies*. Miami, Florida.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13* (5), 431–436.

Gómez, R. L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy, 7* (2), 183–206.

Gout, A., Christophe, A., & Morgan, J. L. (2004). Phonological phrase boundaries constrain lexical access II. Infant data. *Journal of Memory and Language, 51*, 548–567.

Goyet, L., de Schonen, S, & Nazzi, T. (2010). Words and syllables in fluent speech segmentation by French-learning infants: An ERP study. *Brain Research, 1332*, 75–89.

Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science, 18* (3), 254–260.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*, 67–81.

Green, T. R. G. (1979). The necessity of syntax markers: Two experiments with artificial languages. *Journal of Verbal Learning and Verbal Behavior, 18*, 481–496.

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.), *Universals of Language* (pp. 73–113). London, UK: MIT Press.

Guasti, M. T. (2002). *Language acquisition. The growth of grammar.* Cambridge, USA: MIT Press.

Gussenhoven, C, & Jacobs, H. (1998). *Understanding phonology*. London, UK: Arnold.

Harrington, J., Watson, G., & Cooper, M. (1989). Word boundary detection in broad class and phoneme strings. *Computer Speech and Language, 3*, 367–382.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition*, *78*, B53–B64.

Hawkins, J. A., & Gilligan, G. (1988). Prefixing and suffixing universals in relation to basic word order. *Lingua, 74*, 219–259.

Hay, J. S. F., & Diehl, R. L. (2007). Perception of rhythmic grouping: Testing the Iambic/Trochaic Law. *Perception & Psychophysics, 69* (1), 113–122.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies.* Chicago, USA: University of Chicago Press.

Herold, B., Höhle, B., Walch, E., Weber, T., & Obladen, M. (2008). Impaired word stress pattern discrimination in very-low-birthweight infants during the first 6 months of life. *Developmental Medicine & Child Neurology, 50*, 678–683.

Hirsh-Pasek, K., Golinkoff, R., Fletcher, P., DeGaspe-Beaubien, M., & Cauley, K. M. (1985). In the beginning: One word speakers comprehend word order. Paper presented at *Boston University Child Language conference.* Boston, USA.

Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Wright Cassidy, K., Druss, B., Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition, 26*, 269–286.

Hochmann, J.-R., Endress, A. D., & Mehler, J. (2010). Word frequency as a cue for identifying function words in infancy. *Cognition*, *115*, 444–457.

Höhle, B. (2002). *Der Einstieg in die Grammatik: Die Rolle der Phonologie/Syntax-Schnittstelle für Sprachverarbeitung und Spracherwerb.* (Unpublished doctoral dissertation). Freie Universität Berlin, Berlin, Germany.

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborh, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior and Development, 32*, 262–274.

Höhle, B., & Weissenborn, J. (2003). German-learning infants' ability to detect unstressed closed-class elements in continuous speech. *Developmental Science, 6* (2), 122–127.

Hohne, E. A., & Jusczyk, P. W. (1994). Two-month-old infants' sensitivity to allophonic differences. *Perception & Psychophysics, 56* (6), 613–623.

Hualde, J. I. (2003). Segmental phonology. In J. I. Hualde & J. Ortiz de Urbina (Eds.), *A gramar of Basque* (pp. 15–64). Berlin, Germany: Mouton de Gruyter.

Iversen, J., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *Journal of the Acoustical Society of America, 124* (4), 2263–2271.

Jansen, W. (2007). Phonological 'voicing', phonetic voicing, and assimilation in English. *Language Sciences*, *29*, 270–293.

Jessen, M. (1988). *Phonetics and phonology of tense and lax obstruents in German.* Amsterdam, The Netherlands: John Benjamins.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language, 44*, 548–567.

Johnson, E. K., Jusczyk, P. W., Cutler, A., & Norris, D. (2003). Lexical viability constraints on speech segmentation by infants. *Cognitive Psychology, 46*, 65–97.

Johnson, E. K., & Seidl, A. H. (2009). At 11 months, prosody outranks statistics. *Developmental Science, 12* (1), 131–141.

Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science, 13* (2), 339–345.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64,* 675–687.

Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language, 32* (3), 402–420.

Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D. G., Kennedy, L. J., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology, 24,* 252–293.

Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics, 61* (8), 1465–1476.

Jusczyk, P. W., Hohne, E. A., & Mandel, D. R. (1995). Picking up regularities in the sound structure of the native language. In W. Strange (Ed.), *Speech Perception and Linguistic Experience* (pp. 91–119). Baltimore, USA: York Press.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39,* 159–207.

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language, 33* (5), 630–645.

Kahn, D. (1976). *Syllable-based generalizations in English phonology.* (Published doctoral dissertation by Garland Publishing. New York. 1980). MIT, Massachusetts, USA.

Kajikawa, S., Fais, L., Mugitani, R., Werker, J. F., & Amano, S. (2006). Cross-language sensitivity to phonotactic patterns in infants. *Journal of the Acoustical Society of America, 120* (4), 2278–2284.

Kayne, R. (1994). *The antisymmetry of syntax.* Cambridge, USA: MIT Press.

Kehoe, M. M., Lleó, C., & Rakow, M. (2004). Voice onset time in bilingual German-Spanish children. *Bilingualism: Language and Cognition, 7* (1), 71–88.

Kim, J., Davis, C., & Cutler, A. (2008). Perceptual tests of rhythmic similarity: II. Syllable rhythm. *Language and Speech, 51* (4), 343–359.

Kimball, J. P. (1973). Seven principles of surface structure parsing in natural languages. *Cognition, 2*, 15–47.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition, 83*, B35–B42.

Kisilevsky, B. S., Hains, S. M. J., Brown, C. A., Lee, C. T., Cowperthwaite, B., Stutzman, S. S., Swansburg, M. L., Lee, K., Xie, X., Huang, H., Ye, H.-H., Zhang, K., & Wang, Z. (2009). Fetal sensitivity to properties of maternal speech and language. *Infant Behavior and Development*, *32*, 59–71.

Kooijman, V. (2007). *Continuous-speech segmentation at the beginning of language acquisition: electrophysiological evidence*. (Unpublished Doctoral dissertation). Radboud University Nijmegen, Nijmegen, The Netherlands.

Kotz, S. A., Holcomb, P. J., & Osterhout, L. (2008). ERPs reveal comparable syntactic sentence processing in native and non-native readers of English. *Acta Psychologica*, *128*, 514–527.

Kubozono, H. (1989). The mora and syllable structure in Japanese: Evidence from speech errors. *Language and Speech, 32* (3), 249–278.

Kubozono, H. (1996). Speech segmentation and phonological and phonological structure. In T. Otake & A. Cutler (Eds.), *Phonological structure and language processing: Cross-linguistic studies* (pp. 77–94). Berlin, Germany: Mouton de Gruyter.

Kučera, H., & Francis, W. (1967). *Computational analysis of present-day American English.* Providence, USA: Brown University Press.

Ladefoged, P. (1975). *A course in phonetics*. New York, USA: Harcourt Brace Jovanovich.

Ladefoged, P. (2001). *Vowels and consonants.* Oxford, UK: Blackwell.

Langus, A., Marchetto, E., Bion, R. A. H., & Nespor, M. (2012). Can prosody be used to discover hierarchical structure in continuous speech? *Journal of Memory and Language, 66* (1), 285–306.

Langus, A., & Nespor, M. (2010). Cognitive systems struggling for word order. *Cognitive Psychology*, *60*, 291–318.

Lehiste, I. (1960). An acoustic-phonetic study of internal open juncture. *Phonetics, 5* (Supplement 5), 5–54.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20* (3), 384–422.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear & Hearing, 19,* 1–36.

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics, 28* (5), 407–412.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*, 71–102.

Massaro, D. W., & Cohen, M. M. (1983). Phonological context in speech perception. *Perception & Psychophysics, 34* (4), 338–348.

Mattys, S. L. (2004). Stress versus coarticulation: Toward an integrated approach to explicit speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance, 30* (2), 397–408.

Mattys, S. L. & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition, 78*, 91–121.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38*, 465–494.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, *134* (4), 477–500.

Mazuka, R. (1996). Can a grammatical parameter be set before the first word? Prosodic contributions to early setting of a grammatical parameter. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 313–330). Mahwah, USA: Lawrence Erlbaum Associates.

Mazuka, R. (2007). The rhythm-based prosodic bootstrapping hypothesis of early language acquisition: Does it work for learning for all languages? *Journal of the Linguistic Society of Japan, 9* (132), 1–13.

Mazuka, R., & Hayashi, A. (2006). Japanese infants prefer to listen to phonological forms of typical child-directed vocabulary. Paper presented at *15th Biennal International Conference on Infant Studies*. Kyoto, Japan.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language, 39*, 21–46.

McQueen, J. M., Norris, D., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20* (3), 621–638.

Mehler, J., Dommergues, J. Y., Frauenfelder, U., & Seguí, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior, 20*, 298–305.

Mehler, J., Jusczyk, P. W., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29* (2), 143–178.

Mersad, K., & Nazzi, T. (2011). Transitional probabilities and positional frequency phonotactics in a hierarchical model of speech segmentation. *Memory and Cognition, 39*, 1085–1093.

Meuter, R. F. I. & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language, 40*, 25–40.

Morais, J., Content, A., Cary, L., Mehler, J., & Seguí, J. (1989). Syllabic segmentation and literacy. *Language and Cognitive Processes, 4* (1), 57–67.

Morgan, J. L. (1994). Converging measures of speech segmentation in preverbal infants. *Infant Behavior and Development, 17*, 389–403.

Morgan, J. L. (1996). A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language, 35*, 666–688.

Morgan, J. L., & Demuth, K. (1996). *Signal to syntax: An overview*. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 1–22). Mahwah, USA: Lawrence Erlbaum Associates.

Morgan, J. L., Meier, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology, 19*, 498–550.

Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development, 66* (4), 911–936.

Morgan, J. L., Shi, R., & Allopenna, P. (1996). Perceptual bases of rudimentary grammatical categories: Toward a broader conceptualization of bootstrapping. In

J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 263–283). Mahwah, USA: Lawrence Erlbaum Associates.

Mugitani, R., Fais, L., Kajikawa, S., Werker, J. F., & Amano, S. (2007). Age-related changes in sensitivity to native phonotactics in Japanese infants. *Journal of the Acoustical Society of America, 122* (3), 1332–1335.

Mugitani, R., Kobayashi, T., & Amano, S. (2005). Nihongo gakushuu nyuuji no sokuon chikaku tokusei [Japanese learning infants' perception of geminate obstruents]. Poster presented at *Acoustical Society of Japan.*

Nakatani, L. H., & Dukes, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America, 62* (3), 714–719.

Navarro Tomás, T. (1968). *Manual de pronunciación española*. Madrid, Spain: Consejo Superior de Investigaciones Científicas (CSIC).

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, *24* (3), 756–766.

Nazzi, T., Iakimova, G., Bertoncini, J., Frédonie, S., & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language, 54*, 283–299.

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language*, *43*, 1–19.

Nazzi, T., Kemler Nelson, D. G., Jusczyk, P. W., & Jusczyk, A. M. (2000). Six-month-olds' detection of clauses embedded in continuous speech: Effects of prosodic well-formedness. *Infancy, 1* (1), 123–147.

Nespor, M. (2001). About parameters, prominence, and bootstrapping. In E. Dupoux (Ed.), *Language, brain, and cognitive development: Essays in honour of Jacques Mehler* (pp.127–142). Cambridge, USA: MIT Press.

Nespor, M., Guasti, M. T., & Christophe, A. (1996). Selecting word order: the Rhythmic Actication Principle. In U. Kleinhenz (Ed.), *Interfaces in phonology* (pp. 1–26). Berlin, Germany: Akademie Verlag.

Nespor, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H., & Donati, C. (2008). Different phrasal prominence realizations in VO and OV languages. *Lingue e Linguaggio, VII* (2), 1–29.

Nespor, M., & Vogel, I. (1982). Prosodic domains of external sandhi rules. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representations* (pp. 225–255). Dordrecht, The Netherlands: Foris.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht, The Netherlands: Foris.

Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition, 52*, 189–234.

Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21* (5), 1209–1228.

Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition, 83,* B13–B23.

Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language, 53*, 225–237.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language, 32,* 258–278.

Otake, T., Hatano, G. & Yoneyama, K. (1996). Speech segmentation by Japanese listeners. In T. Otake & A. Cutler (Eds.), *Phonological structure and language processing: Cross-linguistic studies* (pp. 183–201). Berlin, Germany: Mouton de Gruyter.

Pallier, C., Sebastián-Gallés, N., Felguera, T., Christophe, A., & Mehler, J. (1993). Attentional allocation within the syllabic structure of spoken words. *Journal of Memory and Language, 32,* 373–389.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural by 8-month-old infants. *Child Development, 80* (3), 674–685.

Peña, M., Bion, R. A. H., & Nespor, M. (2011). How modality specific is the Iambic-Trochaic Law? Evidence from vision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37* (5), 1199–1208.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298,* 604–607.

Peters, A. M. (1985). Language segmentation: Operating principles for the perception and analysis of language. In D. I. Slobin (Ed.), *The crosslinguistic study of* *language acquisition, Vol. 2* (1029–1067). Hillsdale, USA: Lawrence Erlbaum Associates.

Pike, K. L. (1945). *The intonation of American English*. Ann Arbor, USA: University of Michigan Press.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20* (2), 421–435.

Pons, F., & Toro, J. M. (2919). Structural generalizations over consonants and vowels in 11-month-old infants. *Cognition, 116*, 361–367.

Ramus, F. (2002). Language discrimination by newborns: Teasing apart phonotactic, rhythmic, and intonational cues. *Annual Review of Language Acquisition, 2*, 85–115.

Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science, 288* (5464), 349–351.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*, 265–292.

Redington, M., Chater, M., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, *22* (4), 425–469.

Rossi, S., Gugler, M. F., Friederici, A. D. & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience, 18* (12), 2030–2048.

Saffran, J. R. (2001). Words in a sea of sounds: the output of infant statistical learning. *Cognition*, *81*,149–169.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274* (5294), 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27–52.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606–621.

Sanders, L. D., & Neville, H. J. (2000). Lexical, syntactic, and stress-pattern cues for speech segmentation. *Journal of Speech, Language, and Hearing Research, 43*, 1301–1321.

Sanders, L. D., Neville, H., J., & Woldorff, M. G. (2002). Speech segmentation by native and non-native speakers: The use of lexical, syntactic, and stress-pattern cues. *Journal of Speech, Language, and Hearing Research, 45* (3), 519–530.

Santelmann, L. M., & Jusczk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition, 69*, 105–134.

Sato, Y., Sogabe, Y., & Mazuka, R. (2010). Discrimination of phonemic vowel length by Japanese infants. *Developmental Psychology, 46* (1), 106–119.

Scott, D. R. (1982) Duration as a cue to the perception of a phrase boundary. *Journal of the Acoustical Society of America, 71* (4), 996–1007.

Sebastián-Gallés, N., & Bosch, L. (2002). Building phonotactic knowledge in bilinguals: Role of early exposure. *Journal of Experimental Psychology: Human Perception and Performance, 28* (4), 974–989.

Sebastián-Gallés, N., Dupoux, E., Seguí, J., & Mehler, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language, 31*, 18–32.

Selkirk, E. (1978). On prosodic structure and its relation to syntactic structure. In T. Fretheim (Ed.), *Nordic Prosody II* (pp.111–140). Trondheim, Norway: Tapir.

Selkirk, E. (1980). Prosodic domains in phonology: Sanskrit revisited. In M. Aronoff & M.-L. Kean (Eds.), *Juncture* (pp. 107–129). Saratoga, USA: Anma Libri.

Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure.* Cambridge, USA: MIT Press.

Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology Yearbook, 3*, 371–405.

Selkirk, E. (1995). Sentence prosody: Intonation, stress and phrasing. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 550–569). Oxford, UK: Blackwell.

Selkirk, E. (1996). The prosodic structure of function words. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to syntax in early acquisition* (pp. 187–213). Hillsdale, USA: Erlbaum.

Selkirk, E. (2000). The interaction of constraints on prosodic phrasing. In M. Horne (Ed.) *Prosody: Theory and experiments* (pp. 231–261). Dordrecht, The Netherlands: Kluwer .

Shafer, V. L., Shucard, D. W., Shucard, J. L., & Gerken, L. (1998). An electrophysiological study of infants' sensitivity to the sound patterns of English speech. *Journal of Speech, Language, and Hearing Research, 41* (4), 874–886.

Shatzman, K. B., & McQueen, J. M. (2006). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception & Psychophysics, 68* (1), 1–16.

Shi, R., Cutler, A., Werker, J., & Cruickshank, M. (2006). Frequency and form as determinants of functor sensitivity in English-acquiring infants. *Journal of the Acoustical Society of America, 119* (6), EL61–EL67.

Shi, R., & Gauthier, B. (2005). Recognition of function words in 8-month-old French-learning infants. *Journal of the Acoustical Society of America, 117* (4), 2426–2427.

Shi, R., & Lepage, M. (2008). The effect of functional morphemes on word segmentation in preverbal infants. *Developmental Science, 11* (3), 407–413.

Shi, R., Marquis, A., & Gauthier, B. (2006). Segmentation and representation of function words in preverbal French-learning infants. In D. Bamman, T, Mgnitskaia & C. Zaller (Eds.), Proceedings from *BUCLD: 30th Annual Boston University Conference on Language Development, Vol. 2* (pp. 549–560). Boston, USA: Cascadilla Press.

Shi, R., Morgan, J. L., & Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: a cross-linguistic perspective. *Journal of Child Language*, *25* (1), 169–201.

Shi, R., Werker, J. F., & Cutler, A. (2006). Recognition and representation of function words in English-learning infants. *Infancy, 10* (2), 187–198.

Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition, 72*, B11–B21.

Shukla, M., & Nespor, M. (2010). Rhythmic patterns cue word order. In N. Erteschick-Shir & L. Rochman (Eds.), *The sound patterns of syntax* (pp. 174–188). Oxford, UK: Oxford University Press.

Smith, M. R., Cutler, A., Butterfield, S., & Nimmo-Smith, I. (1989). The perception of rhythm and word boundaries in noise-masked speech. *Journal of Speech and Hearing Research, 32*, 912–920.

Smith, R., & Hawkins, S. (2000). Allophonic influences on word-spotting experiments. In A. Cutler, J. M. McQueen and R. Zondervan (Eds.), Proceedings from

Workshop on Spoken Word Access Processes (SWAP) (pp. 139–142). Nijmegen, The Netherlands: Max-Planck-Institute for Psycholinguistics.

Soderstrom, M., Seidl, A., Kemler Nelson, D. G., & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language, 49*, 249–267.

St. Clair, M. C., Monaghan, P., & Ramscar, M. (2009). Relationships between language structure and language learning: The suffixing preference and grammatical categorization. *Cognitive Science, 33*, 1317–1329.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology, 50*, 86–132.

Taylor, D. Q. (1975). The inadequacy of bipolarity and distinctive features: the German "voiced/voiceless" consonants. In P. A. Reich (Ed.), *The Second Lacus Forum* (pp. 107–119). Columbia, USA: Hornbeam Press.

Tabossi, P., Collina, S., Mazzetti, M., & Zoppello, M. (2000). Syllables in the Processing of Spoken Italian. *Journal of Experimental Psychology: Human Perception and Performance, 26* (2), 758–775.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology, 39* (4), 706–716.

Toro, J. M., Nespor, M., Mehler, J., & Bonatti, L. L. (2008). Finding words and rules in a speech stream: Functional differences between vowels and consonants. *Psychological Science, 19* (2), 137–144.

Toro, J. M., Shukla, M., Nespor, M., & Endress, A. D. (2008). The quest for generalizations over consonants: Asymmetries between consonants and vowels are not the by-product of acoustic differences. *Perception and Psychophysics, 70* (8), 1515–1525.

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception & Psychophysics, 67* (5), 867–875.

Trapman, M., & Kager, R. (2009). The acquisition of subset and superset phonotactic knowledge in a second language. *Language Acquisition, 16* (3), 178–221.

Turk, A. E., Jusczyk, P. W., & Gerken, L. (1995). Do English-learning infants use syllable weight to determine stress? *Language and Speech, 38* (2), 143–158.

Valian, V., & Coulson, S. (1988). Anchor points in language learning: The role of marker frequency. *Journal of Memory and Language, 27* (1), 71–86.

van der Lugt, A. H. (2001). The use of sequential probabilities in the segmentation of speech. *Perception & Psychophysics, 63* (5), 811–823.

van Zon, M. (1997). *Speech processing in Dutch: A cross-linguistic approach.* (Unpublished doctoral dissertation). Tilburg University, Tilburg, The Nederlands.

van Zon, M., & de Gelder, B. (1993). Perception of word boundaries by Dutch listeners. In Proceedings from *EUROSPEECH'93: 3rd European Conference on Speech Communication and Technology*, Berlin, Germany.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech, 40* (1), 47–62.

Vroomen, J., & de Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21* (1), 98–108.

Vroomen, J., van Zon, M., & de Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory and Cognition, 24* (6), 744–755.

Warner, N., Kim, J., Davis, C., & Cutler, A. (2005). Use of complex phonological patterns in speech processing: Evidence from Korean. *Journal of Linguistics, 41* (2), 353–387.

Weber, A. (2001). Language-specific listening: The case of phonetic sequences. (Unpublished doctoral dissertation). University of Nijmegen, Nijmegen, The Netherlands.

Weber, A., & Cutler, A. (2006). First-language phonotactics in second-language listening. *Journal of the Acoustical Society of America, 119* (1), 597–607.

Weber-Fox, C. M., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience, 8* (3), 231–256.

Weiss, D. J., Gerfen, C., & Mitchel, A. D. (2009). Speech segmentation in a simulated bilingual environment: A challenge for statistical learning? *Language Learning and Development, 5* (1), 30–49.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology, 24* (5), 672–683.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49–63.

Whalen, D. H., Best, C. T., & Irwin, J. R. (1997). Lexical effects in the perception and production of American English /*p*/ allophones. *Journal of Phonetics, 25*, 501–528.

Woodward, J. Z., & Aslin, R. N. (1990). Segmentation cues in maternal speech to infants. Paper presented at *International Conference on Infant Studies*. Montreal, Canada.

Woodward, J. Z., & Aslin, R. N. (1991). Word-segmentation strategies in maternal speech to infants. Paper presented at *Society for Research in Child Development*. Seattle, USA.

Woodward, J. Z., & Aslin, R. N. (1992). Syntactic and prosodic cues to word segmentation in maternal speech to infants: Turkish and English. Paper presented at *International Society for infant Studies.* Miami, USA.

Yang, C. D. (2004). Universal Grammar, statistics or both? *Trends in Cognitive Sciences, 8* (10), 451–456.

Yip, M. C. W. (2000). Recognition of spoken words in continuous speech: Effects of transitional probability. In B. Yuan, T. Huang & X. Tang (Eds.), Proceedings from *ICSLP'2000* (pp. 758–761). Beijing: China Military Friendship Publish.

Yip, M. C. W. (2006). The role of positional probability in the segmentation of Cantonese speech. In Proceedings from *Interspeech 2006* (pp. 865–868). Pittsburgh, USA: ISCA.

Yoon, Y. B., & Derwing, B. L. (1995). Syllable saliency in the perception of Korean words. In K. Elenius & P. Branderud (Eds.), Proceedings from *13th International Congress of Phonetic Sciences, Vol. 2* (pp. 602–605). Stockholm, Sweden: KTH and Stockholm University.

Yoshida, K. A., Iversen, J. R., Patel, A. D., Mazuka, R., Nito, H., Gervain, J., & Werker, J. (2010). The development of perceptual grouping biases in infancy: A Japanese-English cross-linguistic study. *Cognition, 115,* 356–361.

Zawiszewski, A., Gutiérrez, E., Fernández, B., & Laka, I. (2011). Language distance and nonnative language processing: Evidence from event-related potentials. *Bilingualism: Language and Cognition, 14* (3), 400–411

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, *32*, 25–64.